International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 5351-5371
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

VXLAN/BGP EVPN for Trading: Multicast Scaling Challenges for Trading Colocations

Ashutosh Chandra Jha

Network Security Engineer, New York, USA
*** Corresponding Author Email:** ashutoshjhany@gmail.com - **ORCID:** 0000-0002-7604-9950

Article Info:

DOI: 10.22399/ijcesen.3478
Received : 12 May 2025
Accepted : 14 July 2025

Keywords

VXLAN EVPN
Multicast Scaling
Financial Trading Networks
Head-End Replication (HER)
Market Data Distribution

Abstract:

Financial exchanges today depend on ultra-low-latency market-data delivery, so the underlying network architecture is no longer an afterthought but a strategic asset. In colocation centers where nanoseconds count, VXLAN encapsulation paired with a BGP-based EVPN control plane has become the de facto framework for segmenting and scaling Layer 2 traffic. This paper investigates the role of multicast in that architecture, exposing the technical pressure points network engineers must address to protect the integrity of real-time trading streams. Although EVPN decouples forwarding and forwarding-label assignment, head-end replication still burdens CPU queues on high-density 100-GbE line cards. Multicast is indispensable for distributing top-of-book feeds; however, performance metrics deteriorate sharply once the subscriber count exceeds fifty ports. The study benchmarks the standard draft-based HER against PIM/MBGP and MPLS-VPN, quantifying the round-trip delays, forwarding table churn, and control plane convergence times that each protocol injects into the data path. Experimental findings are paired with field data from production firms, yielding pragmatic recommendations for hardware selection, orderly fan-out patterns, and resilient control-plane adjacency. The paper also reviews nascent tools—segment routing over IPv6, P4 programmable pipelines, and publish-subscribe layers like Apache Kafka—that promise to offload or supplant traditional multicast domains as trading volumes continue their logarithmic growth. By distilling these insights into repeatable design outlines, the article supplies quantitative guidance for architects who must balance performance budgets against regulatory uptime mandates.

1. Introduction

In electronic trading, speed is the only currency that matters. The interval between receiving a tick and placing a hedge can be measured in microseconds, yet carries real monetary weight. Market participants build their systems to exploit every nanosecond; even tiny delays at the network layer distort algorithmic decision-making. To guard against such lags, firms now prefer colocation—data centers positioned within arm's reach of an exchange's matching engine.

Multicast is central to contemporary financial networks. Instead of a one-to-one handshake between sender and receiver, a multicast stream sends data once and allows any device that wishes to listen to pick it up simultaneously. For real-time market content—tick-by-tick prices, order book snapshots, trade reports—this design protects the network by curbing duplicate transmissions and conserving bandwidth. Leading vendors, such as Bloomberg, Nasdaq, and OPRA, have adopted multicast as their distribution backbone, generating packet rates that can peak in the thousands per second and often arrive as sharp bursts. That tempo tests every layer of the underlying hardware. Until recently, firms contained those bursts in Layer 2 clouds, relying on switches to track class-of-service groups with IGMP. Yet, as large financial campuses sprawled and cross-district links grew longer, the weaknesses of flat, broadcast-driven switching became increasingly apparent. Storms of unwanted frames swamped ports, scalability stalled at a few hundred members, and faults spread quickly because isolation boundaries were coarse. Facing these hurdles, trading operations shifted to routed, segment-based designs that offer better visibility, precise bandwidth control, and significantly tighter failure containment.

Trading firms are increasingly relying on VXLAN (Virtual Extensible LAN). This overlay protocol extends Layer 2 networks across a Layer 3 backbone without imposing a near-hard limit on the bridge table size. When combined with BGP EVPN (Border Gateway Protocol Ethernet VPN) as the control plane, VXLAN provides both broad scalability and dynamic learning of MAC and IP addresses. Taken together, these mechanisms enable firms to build wide-area data centers by stitching together Ethernet segments, isolating tenants via virtual routing instances, and updating forwarding tables on the fly. Yet supporting high-bandwidth multicast—especially time-sensitive market feeds—remains a difficult practical question in a VXLAN-BGP EVPN fabric. Although the specification contains hooks for multicast delivery, the original design, favoring unicast, means operators still confront a choice: either replicate packets at every tunnel egress or depend on an underlay multicast scheme that may not scale linearly. Either path risks adding latency, taxing switch CPUs, and ultimately undermining the very performance edge that VXLAN-BGP EVPN is meant to provide.

This paper examines the new multicast challenges that arise in abbreviated trading environments using VXLAN and BGP EVPN. It argues that long-standing multicast techniques must be reconsidered, details the technical shortcomings frequently encountered in live systems, and outlines methods for improving multicast throughput without sacrificing the ultra-low latency that financial firms demand. By pairing practical case studies with laboratory measurements and observations drawn from production networks, the authors offer an honest appraisal of current solutions, their shortcomings, and potential paths forward for multicast in next-generation trading infrastructure.

2. Overview of Trading Colocations and Infrastructure Needs

2.1 Common Architecture in Trading Colocations

Trading colocation facilities are dedicated data centers that house both exchange infrastructure and the servers of competitive trading houses (15). Designing these rooms with minimal long-haul copper or glass runs is crucial because even a few nanoseconds of delay can impact execution quality when markets rush. Firms therefore install bare-metal racks within a few meters of the exchange hardware, linking them with purpose-built, low-jitter fibre. Every circuit board, cable, and switch firmware is specified for deterministic rather than maximum throughput, since variations in latency, rather than absolute speed, introduce risk.

Within that environment, a small yet layered set of systems processes market events. Order gateways serve as the primary ingress and egress points, executing trades with millisecond-level discipline and processing prices and fills in a highly efficient manner. Feed handlers, in parallel, listen to multicast streams from CME, NYSE, or ICE, clean the packets, and timestamp each tick with nanosecond accuracy. The decoded data then flows to analytics boxes, where statistical models, machine-learning kernels, and simple heuristics run against multiple datasets to determine whether to buy, sell, or hedge. Components reside on a flat network fabric, monitored by redundant optical circuits and hot-swap power feeders, allowing hardware failures to be isolated without interrupting the exchange.

As shown in the figure below, the architecture typically involves a linear flow of data from market feed ingestion to analytics and order execution. Each stage plays a critical role in ensuring that trading decisions are made and transmitted with the lowest possible delay, using infrastructure tailored to the exacting requirements of financial networks.

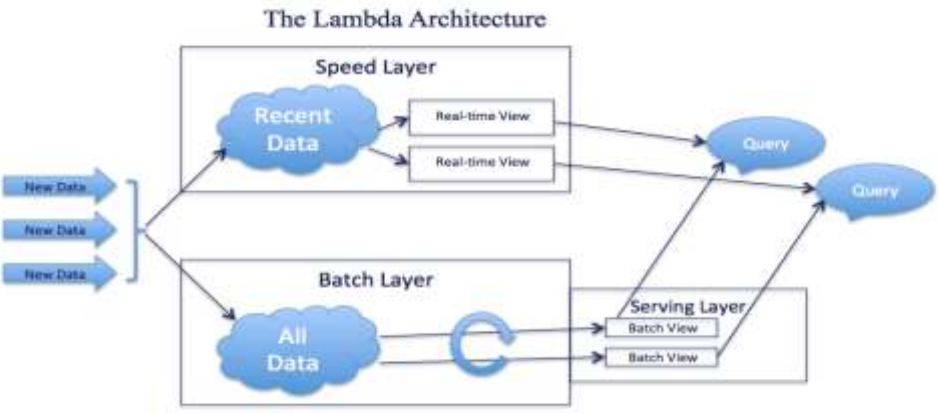

***Figure 1**: Data process flow (source VoltDB)*

2.2 Microsecond-Level Latency Requirements in Equity/Derivatives Trading

Equity and derivatives markets now operate on time scales measured in fractions of a millisecond, with some firms targeting sub-microsecond execution. Within high-frequency trading desks, even a 50-microsecond delay can translate directly into missed liquidity or, worse, material loss on a trade. Achieving this blistering pace requires a network that delivers latency on a deterministic, repeatable basis, completes failover in a time that rounds to zero, and sustains data flow without jitter. Every component along the path—routers, switches, cables, and network interface cards—therefore demands packet-processing code that incurs minimal overhead. Variability introduced by congestion, route flaps, or inefficient packet replication, however slight, quickly becomes a risk that the trading algorithm cannot tolerate.

2.3 Infrastructure Evolution: From Flat Layer 2 to EVPN-Based Overlays

Market participants for years have built trading networks around flat Layer 2 topologies, which permit multicast propagation and eliminate the need for time-consuming routing (35). That simplicity helped early firms engineer connectivity quickly within small trading halls. Growth, however, brought complexity: collocated venues added tenants, message feeds swelled, and applications diversified across multiple operating profiles. Under these conditions, classic broadcast domains invited storms, spanning-tree recalculations lengthened convergence windows, and MAC-address conflicts polluted forwarding tables. The cumulative operational noise slowed provisioning, increased troubleshooting effort, and capped scalability. To counter these issues, many exchanges and quant firms migrated toward EVPN-based overlays, which confine broadcast traffic, expedite convergence, and isolate tenant workloads within a single, logical fabric—reflecting broader principles of architectural diversification and scalability seen in infrastructure planning strategies (18). To address these constraints, organizations increasingly implement VXLAN (Virtual Extensible LAN) as an overlay mechanism that extends Layer 2 broadcasts across a conventional Layer 3 backbone. Pairing this with BGP EVPN as the control plane delivers automatic address learning, destination resolution, and tenant boundary enforcement. Consequently, multiple trading firms, applications, and services can co-exist on the same hardware while remaining isolated in discrete virtual segments. The combined solution also enhances scalability, operational efficiency, and fault containment throughout the data center fabric.

As the table below illustrates, EVPN overlays offer clear advantages in scalability, performance, and operational manageability:

***Table 1**: Comparison of Flat Layer 2 Networks and EVPN-Based Overlays in Trading Infrastructure*

Aspect	**Flat Layer 2 Networks**	**EVPN-Based Overlays (with VXLAN)**
Multicast Support	Native, simple propagation	Controlled, tenant-aware distribution via HER or native multicast
Network Scalability	Limited due to broadcast domain size and MAC table growth	High scalability with Layer 3 backbone and address learning via EVPN
Tenant Isolation	Minimal; prone to MAC/address conflicts	Strong isolation through VNIs and EVPN route control
Fault Containment	Weak; broadcast storms affect the entire domain	Improved through segmentation and confined broadcast domains
Convergence Time	Slow due to STP recalculations	Fast convergence with BGP EVPN and Layer 3 core
Operational Complexity	Grows with size; manual MAC provisioning and STP tuning	Reduced via automated MAC/IP learning and centralized policy enforcement
Provisioning Speed	Slower as scale increases	Faster deployment using EVPN route distribution and automation
Support for Multiple Firms	Risk of overlap and leakage	Secure multitenancy via logical segmentation (VNIs/VRFs)

2.4 Role of Redundancy, Segmentation, and Real-Time Telemetry

Redundancy and logical segmentation are essential for maintaining uptime and data integrity in mission-critical trading systems. Strategically placed redundant links, dual-homed switches, and active-active paths minimize both outright outages and the latency of planned failover. Meanwhile, network segments identified by unique virtual network identifiers (VNIs) ensure that distinct strategies, teams, or clients remain logically separate yet receive the same exchange feeds in real time and without packet interference. Real-time telemetry has become a cornerstone of colocation facilities (23). Live readings of packet flow, jitter, and congestion enable network managers to identify irregular patterns before they mature into visible service degradation. These telemetry feeds are commonly ingested by centralized monitoring platforms, which in turn trigger automated measures such as dynamic traffic rerouting or on-the-fly adjustment of quality-of-service codes. In trading halls, where even a few extra milliseconds during a volatile event can translate directly into revenue loss, that level of immediate insight is indispensable. These capabilities, along with others already discussed, constitute the architectural bedrock of contemporary electronic trading venues. The transition from conventional Layer-2 fabrics to VXLAN/BGP EVPN overlays addresses a pressing need for scalable, logically isolated, and fault-tolerant networks that can accommodate the extreme workload fluctuations characteristic of financial markets.

3. VXLAN and BGP EVPN: Technical Primer

3.1 VXLAN Explained: Overlays, Tunnels, VTEPs, VNIs

VXLAN, or Virtual Extensible LAN, is an overlay technique crafted specifically to push past the scalability ceiling that shadows conventional VLAN setups. In a classic Layer-2 data-center architecture, VLANs carve up traffic neatly; however, the 4096-tag cap imposed by the 12-bit identifier quickly becomes a bottleneck in places such as trading colocation racks, where dozens of isolated segments must coexist side by side. The VXLAN approach strips that ceiling. By sitting on top of an already running Layer-3 backbone, it builds virtual Layer-2 domains through encapsulation: plain Ethernet frames are encapsulated into UDP packets and travel across the routable fabric. That wrapping takes place at the network edge inside a VXLAN Tunnel Endpoint, which may reside in a physical switch or a software-based virtual appliance. Every new segment receives a 24-bit VXLAN Network Identifier (VNI), which provides the system with room for approximately 16 million tunnels —a number that dwarfs the VLAN footprint. For high-frequency trading floors and multi-tenant clouds, VXLAN provides horizontally scalable, software-defined Layer 2 lanes that allow dozens of applications or customer pods to run in strict isolation while sharing standard wires.

As shown in the figure below, VXLAN builds a logical Layer 2 overlay across an IP-based underlay by encapsulating and routing packets between VTEPs, each associated with a unique VNI segment.

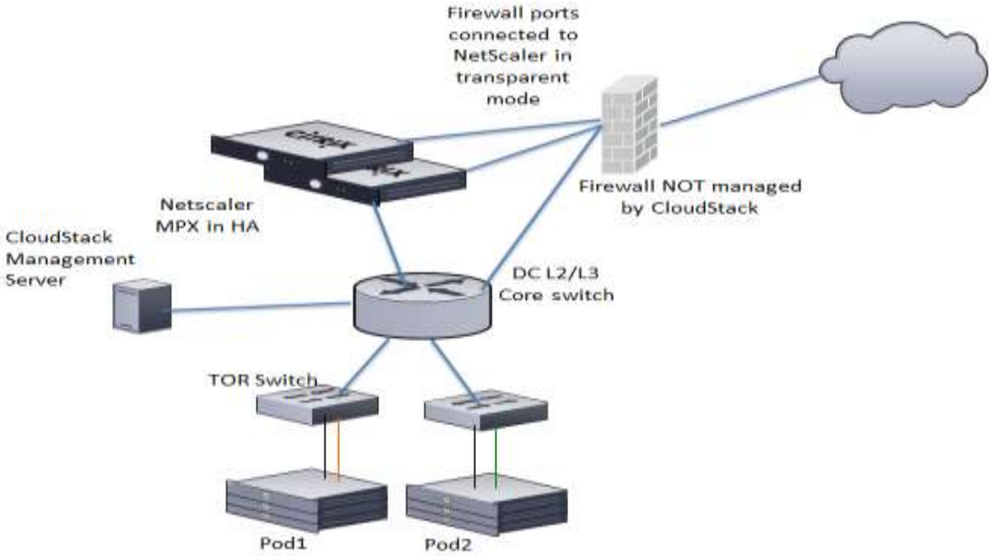
Firewall ports connected to NetScaler in transparent mode
Netscaler MPX in HA
Firewall NOT managed by CloudStack
CloudStack Management Server
DC L2/L3 Core switch
TOR Switch
Pod1
Pod2

Figure 2*: networking_and_traffic*

3.2 BGP EVPN as the Control Plane: Route Types 2, 5, and 6

Although VXLAN specifies the encapsulation and forwarding mechanisms for overlay packets, it does not provide a systematic way for devices to learn where MAC or IP addresses reside. That function falls to BGP EVPN, which serves as the control plane for VXLAN segments. Rather than forward packets unquestioningly until they find a sink, EVPN shares address bindings via structured, incremental route announcements—a process that enhances data consistency and routing precision across distributed systems, much like consistency mechanisms employed in distributed databases such as MongoDB (12, 13). Three route types are particularly noteworthy. Route Type 2 carries MAC address bindings and, optionally, overlay IP addresses, allowing remote VTEPs to load their forwarding tables with minimal reliance on flooding. Route Type 5 advertises IP prefixes, thus enabling inter-subnet routing within the extended broadcast domain created by the overlay. Finally, Route Type 6 forwards multicast group information, letting the fabric intelligently steer market data and price streams only to interested peers (20). Collectively, these mechanisms trim unnecessary control-plane chatter, lighten data-plane flooding, and accelerate convergence—attributes that matter in any network but are indispensable when millisecond latency and deterministic packet delivery define service quality.

3.3 Benefits for Trading Setups: Scale, Segmentation, Dynamic Learning

Integrating VXLAN with BGP EVPN introduces a suite of operational advantages for trading-floor networks. Principal among them is scale. By providing millions of available VNIs, firms can allocate discrete overlay segments to individual desks, strategies, or even external clients, eliminating identifier clashes through systematic prefix management. Segmentation follows as a natural extension. Each team or business unit receives its logical Layer 2 domain, allowing for strict policy enforcement while still utilizing shared hardware. This separation enhances data privacy and supports auditors by establishing clearly defined boundaries that meet international regulatory standards with minimal additional tooling.

Dynamic MAC and IP learning delivered via BGP further sharpens responsiveness. Network nodes advertise learned addresses instead of flooding them across the fabric, cutting broadcast noise and accelerating convergence after link failures or topology changes. The combination is especially valuable in volatile markets, where microsecond delays can affect trading outcomes.

As illustrated in Table 2 below, the adoption of VXLAN EVPN introduces tangible benefits that directly align with the performance, scalability, and compliance demands of financial trading environments:

***Table 2**: Key Benefits of VXLAN EVPN in Trading Network Setups*

Feature	**Flat Layer 2**	**VXLAN EVPN**
Scalability	Limited to VLAN ID space (≈4K)	Supports over 16 million VNIs for granular segmentation
Segmentation	Coarse separation; risk of overlap	Per-strategy or per-client overlays with full isolation
MAC/IP Learning	Flood-based learning with high broadcast traffic	Control-plane learning via BGP, reducing noise and improving stability
Compliance Support	Manual zoning and complex ACLs	Logical separation supports auditing and regulatory reporting
Convergence Speed	Slow during topology changes	Fast recovery through BGP-triggered updates

3.4 Comparison to Legacy L2VPN/MPLS Deployments

Before VXLAN/EVPN, many firms built Layer 2 overlays on MPLS, leaning heavily on VPLS or dedicated circuits. Such designs provided predictable latency, yet they were brittle. Adding a new circuit or isolated service entailed manual tweaks on every intervening label-switch router, limiting agility and tying capacity to the underlying physical mesh. Because these constraints emerge quickly in an expanding trading operation, administrators often find themselves choosing between additional hardware and the risk of oversubscribed links (2). In contrast, the EVPN control plane abstracts provisioning with uniform route advertisements, allowing operators to scale horizontally for modest incremental cost while retaining end-to-end latency characteristics that meet institutional requirements. VXLAN overlay paired with BGP-based EVPN significantly updates traditional L2-L3 interconnection methods. It automates provisioning, learns endpoints dynamically, and supports multiple tenants natively, all atop standard IP gears. This evolution delivers quicker deployment, clearer visibility, and stronger fault tolerance—qualities traders now expect to stay competitive.

4. Importance of Multicast in Trading Networks

4.1 Core Use Cases: Market Data Distribution, Order Books, Reference Data

In finance, market data is the most urgent flow crossing the network. Live quotes, bids, offers, trades, options, futures, and indices arrive in steady, time-critical streams. Vendors like Bloomberg, Nasdaq, ICE, and OPRA single-hop the feeds from central servers to hundreds, sometimes thousands, of traders located in colocation facilities. Each feed pulses high-volume, high-frequency updates that can tip a real-time trading decision in milliseconds. In addition to streaming price ticks, multicast carries the complete order-book picture, showing traders how many shares sit at each level and how those levels shift over time. Algorithms watch this evolving depth to gauge available liquidity, spot emerging patterns, and place their trades milliseconds ahead of small price changes. Critical reference data—corporate actions, symbol mappings, trading halts—also rides the same multicast streams so that every workstation starts with the same baseline facts.

As illustrated in the figure below, these distinct streams—live pricing, order depth, and reference data—are closely interlinked. They collectively drive the informational backbone that powers both manual trading desks and automated execution engines in colocation environments.

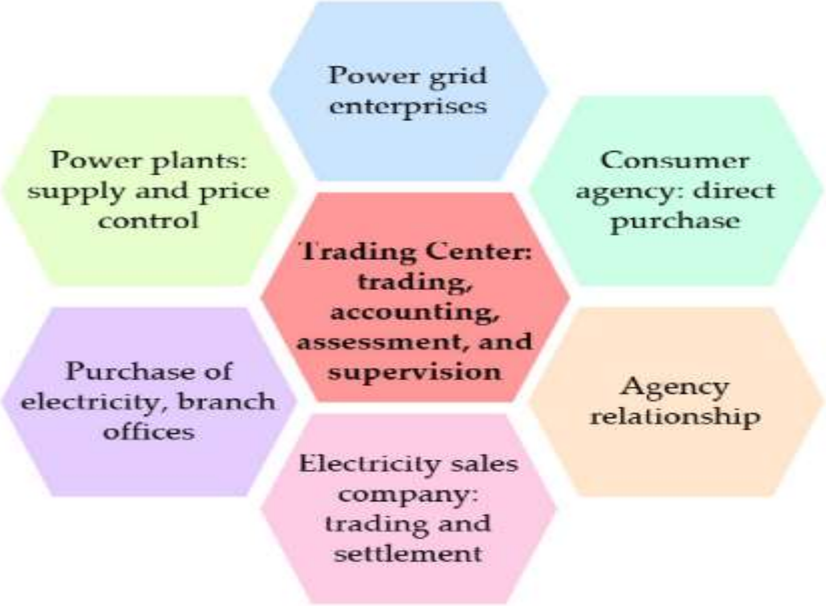
Power grid enterprises
Power plants: supply and price control
Consumer agency: direct purchase
Trading Center: trading, accounting, assessment, and supervision
Purchase of electricity, branch offices
Agency relationship
Electricity sales company: trading and settlement

Figure 3: *Main relationships between the respective markets.*

4.2 Why Multicast is Favored Over Unicast: Scalability, Latency, Efficiency

That shared delivery makes multicast the apparent choice for high-volume feeds. If the exchange switched to unicast, the sender would waste bandwidth repeating the duplicate packets for each client, burden the network with duplicate traffic, and overload its processing pipelines. By contrast, multicast pushes a single copy, frees precious router resources, and scales gracefully as new receivers join without taxing the stream's original sender. One broadcast stream is sent once and then routed to anyone who wants it. In trading rooms where dozens or hundreds of systems—such as feed handlers, risk engines, backup nodes, and others—must receive the same tick simultaneously, the single-send method scales better than separate copies. It also reduces latency, as devices do not wait in line for their turn, and lightens the bandwidth load, a relief during peak trading hours when packets surge through the pipe. Within a financial colocation cage, that design means every trading server receives a fresh update with minimal jitter, maintaining a competitive edge and providing all operators with a roughly equal, predictable window to act.

4.3 Behavior Patterns: Short-Lived Bursts, High Message Rate (e.g., 5k pps)

Market data feeds do not pour out a steady stream; instead, they roar during short, news-driven periods. Across the first few ticks of market open, the final pause before close, or the moment an economic release hits the wire, packet counts can leap past 5000 per second on a single channel, and a sizeable trading firm might track dozens of those channels all at once. Multicast delivery in Layer 2 colocation networks begins with switches monitoring Internet Group Management Protocol (IGMP) messages, a feature known as IGMP snooping. By observing which hosts join or leave particular multicast groups, switches can build membership tables that guide traffic only to relevant ports. This selective forwarding prevents multicast packets from being broadcast throughout the entire domain, reducing unnecessary load on already busy links and preserving bandwidth for unicast and broadcast frames. In dense racks where thousands of servers share the same layer, throttling flood traffic is critical for protecting microsecond-scale latencies demanded by latency-sensitive applications—a requirement increasingly recognized in both networking and real-time systems, where timing precision underpins performance and reliability (21, 28).

Devices indicate their wish to join specific multicast groups using messages defined by IGMP, the Internet Group Management Protocol. Accordingly, switches eavesdrop on these announcements and update their forwarding tables so that only receivers listed in the table see the corresponding traffic. Because the Layer 2 fabric processes multicast in hardware with minimal additional configuration, it can deliver a straightforward and high-speed solution for networks that prioritize operational simplicity. As infrastructures expand and organizations require multitenancy, clear traffic separation, and room to grow, the limitations of pure Layer 2 multicast become apparent. In flat broadcast domains, isolating traffic, mitigating storm events, and enforcing tenant-specific policies remain challenging, which can lead to congestion and service disruptions. These shortcomings prompt architects to investigate multicast within VXLAN-BGP-EVPN designs, where layered mechanisms can maintain high performance while satisfying demands for flexibility and control.

5. Multicast over Traditional vs. Modern Architectures

5.1 Traditional L2 Multicast: IGMP Snooping, STP, Flooding Issues

In conventional Layer 2 Ethernet networks, multicast packets are primarily handled through Internet Group Management Protocol (IGMP) snooping, combined with Spanning Tree Protocol (STP). By listening to IGMP join and leave messages, a switch learns which ports require specific multicast streams and forwards those packets only to the relevant links, thus preventing unnecessary flooding throughout the entire LAN. Although this technique performs reasonably well in small, relatively static environments, it reveals apparent weaknesses when deployed in trading colocation facilities (7). First, IGMP snooping relies on timely and accurate membership updates. During brief, high-rate multicast surges, the state table can converge too slowly, resulting in packets flooding anyway. Second, STPs' loop-prevention mechanism lengthens failover times, potentially resulting in several seconds of lost time. At the same time, the network waits for blocked links to move to the forwarding state, a delay that directly translates into missed market data.

In dense colocation facilities housing dozens or hundreds of multicast streams and supporting microsecond-level price updates, the old Layer 2 model quickly becomes brittle. Membership churn, transient equipment malfunctions, or even minor configuration drift can trigger flooding, packet duplication, or loss—behaviors that undermine latency-sensitive applications and strain already limited switch buffers. These cascading effects mirror challenges found in distributed microservice ecosystems, where loose coupling and event-driven patterns demand strict control of timing and communication boundaries to maintain stability (9, 10).

5.2 Overlay Multicast Expectations: Tenant-Aware Replication, Elasticity

Contemporary overlay architectures, implemented via VXLAN and BGP-EVPN, confront traditional multicast hurdles by providing tenant-conscious and elastic networking. Within this framework, overlay multicast must accommodate fluid group membership across geographically dispersed virtual segments, confine traffic strictly within tenant boundaries, and expand horizontally across data-center pods with minimal, ideally automated, configuration. Central to this goal is tenant-aware replication, which is mandated to honor virtual network identifiers (VNIs), ensuring that multicast packets generated in one tenant domain remain isolated from all others. The architecture must likewise demonstrate elasticity, accommodating rapid changes in group cardinality or receiver counts while upholding low latency and high reliability thresholds. Realizing these requirements is non-trivial, partly because VXLANs were initially optimized for unicast payloads. Implementing Overlay multicast therefore demands supplementary protocols or enhancements that either mimic traditional multicast behaviors or natively support them within the VNI-scoped tunneling model.

5.3 Head-End Replication (HER): Simplicity versus Performance Trade-Offs

Head-End Replication (HER) is the predominant mechanism employed for multicast traffic within VXLAN overlay networks. Under this approach, the originating VTEP duplicates each multicast packet and transmits it separately to every remote VTEP with interested receivers. By offloading the multicast routing burden from the physical Underlay, HER simplifies deployment in Layer-3 spine-leaf architectures where native multicast treatment is either unsupported or administratively undesirable. From a practical engineering standpoint, HER is straightforward to configure and bypasses the management overhead associated with Protocol Independent Multicast (PIM) and its control-plane signaling. Nevertheless, this convenience carries a marked performance penalty. Because packet copying occurs at the ingress VTEP, the forwarding hardware must simultaneously process and transmit multiple duplicates of the same frame. Such workload demands substantial CPU cycles and consumes aggregate bandwidth, particularly when hundreds of receivers subscribe to one or more active groups. In high-density trading environments, traffic can surge to several gigabits per second during scheduled events, such as market open, placing additional strain on the HER pipeline. Under these conditions, replication may saturate the sending chassis, introducing jitter, packet loss, and consequential delivery delays—metrics that in a financial context render the service unusable.

5.4 Native Multicast: PIM-Based Models and Pruning Behavior

One alternative to Hybrid Edge Replication (HER) is to rely on native multicast within the Underlay and build traffic trees using Protocol Independent Multicast—Sparse Mode (PIM-SM) (26). By constructing these trees at the underlay layer, packets are forwarded only along links that have active receivers, thereby pruning excess traffic and curtailing duplicate copies. This pruning mechanism is particularly valuable in environments that carry large data streams, where spare bandwidth should be conserved. When compared with HER, native multicast scales more gracefully in settings featuring numerous groups or widely dispersed listeners. Offloading replication duties to PIM-capable routers lightens the processing burden on the source virtual tunnel endpoint, and hardware-forwarding engines in core switches propagate traffic with minimal CPU intervention.

The PIM-based design is not without hurdles, however. Network operators must configure multicast-routing protocols, assign and manage Rendezvous Points (RPs), and ensure that overlay signaling and underlay state remain in sync. Moreover, many data center teams choose to turn off underlay multicast due to perceived operational overhead or because merchant silicon switches still lack robust support for the feature. In real-world trading networks, some firms opt for a hybrid model, applying head-end replication for small multicast groups and relying on native Protocol-Independent Multicast (PIM) for large, high-throughput feeds. The final decision usually hinges on receiver count, replication load, available hardware, and the organization's appetite for added operational complexity.

As shown in the figure below, PIM constructs efficient tree-based distribution paths, which contrast sharply with HER's replicated point-to-point approach.

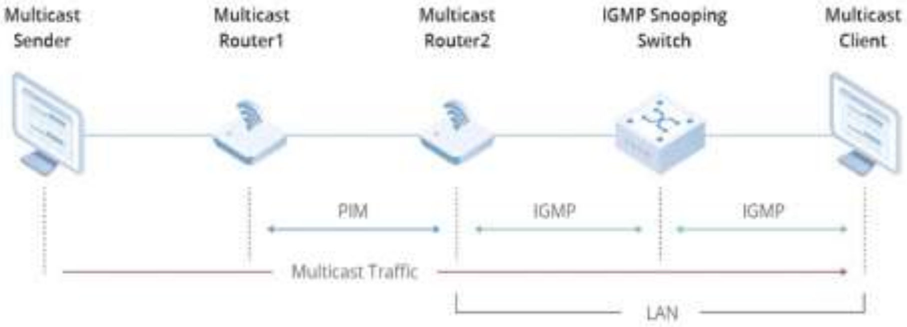

***Figure 4**: Protocol Independent Multicast*

6. Multicast Scaling Challenges in VXLAN/BGP EVPN

The combination of VXLAN and BGP-based Ethernet Virtual Private Networks undoubtedly adds flexibility to trading architectures; however, it also presents scaling challenges for multicast delivery. Such problems surface quickly in high-frequency settings, where even microsecond delays or dropped clones can have a significant impact on the market. Much like the trade-offs observed in dynamic inference systems—where timing and structural design influence accuracy and performance—the reliability of multicast in EVPN overlays hinges on tight coordination and resource efficiency (27).

6.1 Replication Pressure on VTEPs in HER Scenarios

In many VXLAN-EVPN rollouts, especially those deliberately avoiding native multicast in the Underlay, firms default to head-end replication for multicast transport. Under this scheme, the originating tunnel-end-point duplicates each packet toward every downstream VTEP that advertises interest in it. That pattern holds firm with small groups, but once the number of receivers reaches the dozens or hundreds, each new addition adds a fresh copy, and the source node bears a replication load that climbs linearly.

In trading colocation facilities where a single market-data multicast is delivered to dozens of servers spread across numerous racks or even distinct pods, the originating switch can experience severe CPU and memory strain. That load intensifies whenever message rates spike; at thousands of packets per second, the switch must simultaneously parse, replicate, and forward each frame without introducing unacceptable latency. If the virtual tunnel endpoints (VTEPs) supporting that traffic are not sized for high-throughput duplication, they may drop packets, throttle replication queues, or inject jitter—all of which compromise the integrity of the market feed.

6.2 Feed Storm Impact during Market Open/Close

The trading day begins and ends with its highest stress. At market open and again at the close, incoming bids, offers, and last-sale prints surge, forcing network pipes into temporarily sustained overloads. This phenomenon, popularly termed a feed storm, appears as a sharp spike in both the rate and total volume of multicast traffic coursing through the data center fabric. Storms of burst traffic arriving from outside the overlay can overwhelm the VTEPs that act as ingress points. In a HER-based VXLAN EVPN fabric, this surge forces the original edge hardware to replicate packets for every reachable peer before handing them to the traditional forwarding plane. That demand places pressure on lookup pipelines, memory queues, and link bandwidth simultaneously. When the combined packet rate pedals beyond the VTEPs' processing envelope, the result is pronounced latency spikes, buffer overrun events, and—under heavy strain—packet loss. Such degradation is intolerable in high-frequency trading contexts, where microsecond accuracy is built into profit and risk models; even a single tick of latency or an errant dropped frame can distort edge algorithms, trigger erroneous orders, or forfeit timely arbitrage opportunities.

6.3 Absence of Native Group Pruning in HER

HER further compounds this strain by forgoing native multicast pruning. Under current configurations, every VTEP that subscribes to a given VNI receives the complete fan-out of multicast traffic, regardless of whether any downstream endpoint has expressed interest in a particular group. Consequently, packets arrive at switches and link segments that lack active consumers, wasting output bandwidth and elevating jitter on already-stressed interconnects. HER's handling stands in marked contrast to mature IP multicast schemes, which utilize Protocol Independent Multicast to prune inactive branches and confine replication to paths verified on demand. By omitting such selective duplication, HER delivers a coarse, catch-all model that escalates overhead, degrades average packet delivery time, and fragments forwarding capacity—weaknesses that are acutely visible in multi-tenant clouds or large production fabrics where every byte counts.

6.4 Complexity in Deploying PIM with EVPN in the Underlay

Although native multicast alleviates several scalability hurdles that arise in Hybrid Edge Routers, embedding it within an EVPN-based data center presents exacting technical demands. Chief among these is the requirement to activate Protocol Independent Multicast (PIM) throughout the Underlay, introducing an additional layer of routing logic that operators must now consider. The predominant implementation, PIM Sparse Mode (PIM-SM), relies on Rendezvous Points, shared trees, and dynamically built source trees; each of these components must be carefully designed, documented, and maintained over the network's operational lifetime. This mirrors challenges seen in predictive analytics systems, where layered complexity must be managed continuously to avoid performance degradation and ensure operational efficiency (22). The task becomes even more nuanced when PIM is integrated with BGP-driven EVPN. Consistent forwarding tables, mirrored control-plane logic, and synchronous group membership state must be propagated across all devices; any departure from this state can result in silent data loss or wasted forwarding capacity. Because the architecture is so state-sensitive, many production teams regard it as brittle unless high-grade change-management policies and frequent health checks are applied. Furthermore, implementation scope may be curtailed by hardware constraints: not every switch vendor permits full PIM functionality and VXLAN-EVPN to coexist on the same forwarding plane, which in turn limits topological elasticity and forces operators to reconsider vendor selection early in the design phase.

6.5 Group Collision and Network Isolation among Tenants

In a multi-tenant trading colocation facility, diverse financial firms often provision identical multicast addresses to aggregate market feeds and real-time executions across their internal stacks (8). Under a conventional Layer-2 architecture, that overlap can result in unintended traffic cross-pollination between logical tenants, threatening data confidentiality as well as sustaining predictable control-plane performance. Migrating to an EVPN fabric nominally addresses this risk by binding multicast intents to specific Ethernet segments; however, providers must still exercise disciplined address management and deploy appropriate filtering rules to enforce strict policy separation on every tropical hop.

VXLAN over BGP EVPN utilizes VNIs and VRFs to segment tenant traffic; however, HER-style replication groups do not automatically confine multicast packets to their intended subscribers. Absent a stringent overlay-level scoping of group-management policies, packets destined for one tenant may inadvertently traverse the VTEPs of another, exposing sensitive data to unintended recipients. In regulated financial networks, such leakage is a non-negotiable risk that governing bodies will scrutinize. Achieving the necessary level of isolation, therefore, demands disciplined alignment of address prefixes, route-target communities, and per-VNI group filters—tasks that HER does not provide out of the box. Designers thus encounter an additional layer of administrative overhead, requiring detailed documentation, consistent policy audits, and validation testing to assure that multicast boundaries remain impermeable across the data plane.

7. BGP EVPN and Multicast Integration Techniques

VXLAN EVPN was initially tailored for unicast traffic; however, the high-speed requirements of trading systems have prompted network designers to seek reliable, low-latency multicast solutions. Supporting multicast in an EVPN fabric, therefore, requires supplementary control-plane signaling, careful integration with legacy multicast protocols, and thorough testing against platform-specific ASIC capabilities. Just as logistics operations rely on precision-driven, algorithmic coordination to meet stringent delivery timelines, EVPN multicast deployments demand exacting architectural alignment and operational discipline. The following subsections survey the leading techniques and prevailing difficulties that operators encounter when merging multicast services into a BGP-based EVPN overlay. (25).

7.1 Multicast with EVPN: Route Type 6 and Signaling Extensions

To natively circulate multicast membership information within the VXLAN control plane, EVPN now defines Route Type 6, the Inclusive Multicast Ethernet Tag (IMET) route. By encoding the multicast group address and associated VNI in a single BGP update, a forwarding instance can register interest in group traffic without deferring to IGMP or PIM. This adjustment minimizes elegant flooding while allowing policy-based replication decisions to mature progressively at the fabric edge. When a VTEP wishes to join a multicast class, it issues an IMET announcement listing the target group and the corresponding VNI. Other VTEPs collect these routes—whether operating ingress replication, sparse-mode PIM, or switched-plane native multicast—and apply the advertised memberships to either local replication lists or tree-joining processes. The approach thus combines traditional multicast semantics with VXLAN flexibility, although inter-layer consistency and hardware resource contention remain key operational concerns. Recent IETF drafts also introduce signaling enhancements that link EVPN with legacy multicast domains, blending source-group (S, G) and shared-group (S, G) semantics into the EVPN control plane. These additions provide operators with stronger visibility and finer control over group memberships across the VXLAN overlay. Table 3 below summarizes the functional aspects of Route Type 6 and related EVPN multicast signaling methods:

***Table 3**: Functional Role of Route Type 6 and EVPN Multicast Signaling*

Aspect	**Route Type 6 (IMET)**	**Signaling Enhancements**
Purpose	Announce multicast group interest in EVPN overlay	Integrate (S, G) and (*, G) models into BGP EVPN signaling
Use Case	Populate VTEP multicast membership tables	Enable source-aware multicast across overlays and legacy underlays
Dependency on PIM/IGMP	Eliminated for intra-EVPN multicast signaling	Compatible with PIM/IGMP for hybrid environments
Replication Model Compatibility	Works with HER, PIM-SM, or hardware-native multicast	Supports seamless transition between overlay and underlay multicast
Operator Benefit	Reduced flooding, improved scale and policy control	Enhanced visibility and multicast control in complex data center fabrics

7.2 Ingress Replication (HER) vs. Native Multicast over Underlay (PIM-SM, MVPN)

Two primary approaches deliver multicast in VXLAN EVPN fabrics: ingress replication, often referred to as HER, and native multicast routed through underlay protocols such as PIM-SM or

MVPN. Ingress replication is conceptually straightforward. When a packet reaches a sending VTEP, that node duplicates it for each peer listed in its IMET and transmits those copies individually, because the underlay does not require multicast routing state, operators can quickly deploy the feature. Yet, high membership counts and congested packet streams can overload CPU and bandwidth on the source node.

Native multicast shifts the replication burden to the underlay (33). Routers use Protocol Independent Multicast-Sparse Mode to assemble a shared tree, and the VTEP joins the tree rather than forwarding multiple duplicates. That architecture prunes branches without receivers, preserving link capacity, lowering processing cycles on the source VTEP, and scaling better in environments where multicast traffic patterns are unpredictable. (29, 30). A more sophisticated approach to integrating multicast within an Ethernet VPN fabric employs Multicast Virtual Private Network techniques, distributing multicast routing information across Border Gateway Protocol and encapsulating it according to RFC 6514. This method enables administrators to exert detailed control over which multicast groups are propagated throughout the network, while enhancing compatibility with existing EVPN control-plane messages. However, realizing these advantages typically demands extensive device configuration and reliable vendor support across all network nodes.

7.3 Vendor Implementation Status and Differences

Multicast support within EVPN fabrics varies significantly across major vendor platforms in terms of both feature scope and operational stability. Cisco's Nexus line presents arguably the most comprehensive toolkit, including head-end replication (HER), Protocol Independent Multicast (PIM), Route Type 6 signaling, and optional anycast rendezvous points. Supplementary capabilities, such as selective replication and improved processing of broadcast, unknown unicast, and multicast (BUM) frames, enable data-center operators to optimize bandwidth usage and maintain predictable low latency in high-density colocation environments.

Arista switches, by contrast, default to HER for multicast traffic throughout the fabric. Although PIM-based native multicast is supported, it is confined to selected hardware models and only functions under narrowly defined configuration conditions. Many Arista platforms also limit the scale of multicast streams and the underlying IP fabric that can be leveraged. Consequently, network architects planning low-latency use cases—especially in trading floors or media distribution—must carefully match workload requirements to the intended switch series early in the design cycle, thus assuring that replication and delivery goals will be consistently met. Juniper distributes its multicast EVPN features across the QFX and MX families, carrying traffic in both hardware and software forms. The routers and switches handle either the hierarchical EVPN (HER) model or the more sophisticated MVPN scheme, with implementation rigor aimed at meeting recognized standards and working uniformly across different vendors. Because the design aligns with current RFCs as well as experimental drafts, the system can be easily integrated into multi-supplier data centers without requiring major adjustments.

Real-world tests still reveal minor mismatches in forwarding logic. Such mismatches become noticeable whenever a hardware-based VTEP interacts with a software-based router or when operators move from HER to pure native multicast. Because trading workloads tolerate little jitter or added latency, thorough pre-deployment performance checks, including frame capture and delay measurement, should guide the final production decision. As shown in the figure below, migrating traditional Ethernet fabrics to VXLAN EVPN requires not only architectural adaptation but also careful platform selection and replication-mode validation to ensure consistent performance.

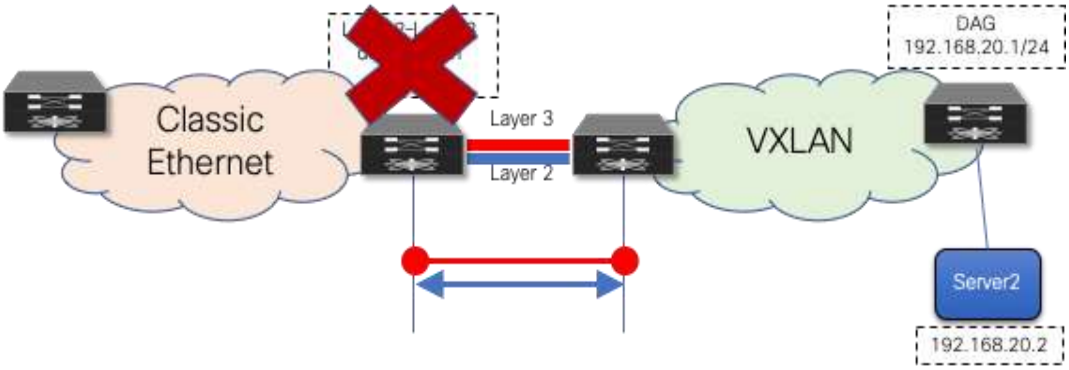

***Figure 5**: Migrating Classic Ethernet Environments to VXLAN BGP EVPN*

7.4 Integration of EVPN with MVPN (Drafts, RFCs, Practical Viability)

The integration of MVPN with EVPN is detailed in ongoing IETF work, particularly in the draft titled *draft-ietf-bess-evpn-mcast*. This draft outlines mechanisms for enabling multicast services within EVPN fabrics by leveraging the principles of Multicast VPNs. The approach builds upon foundational standards, most notably RFC 6513 and RFC 6514, which define the architecture and signaling procedures for multicast in BGP/MPLS VPN environments. Together, these documents provide a framework for scalable, efficient multicast delivery across EVPN overlays using established MVPN techniques.

When deployed together, EVPN and MVPN enable source-specific multicast (SSM) and shared-tree architectures to run within an overlay, allowing for targeted group pruning, selective replication, and fast failover. In reality, though, roll-out has lagged. Administrators cite heavy configuration overhead, the demand for feature-rich hardware, and patchy vendor interoperability as ongoing barriers to widespread use.

In trading colocation facilities, where ease of setup, predictable behavior, and near-instant convergence matter most, hop-exit replication (HER) remains the preferred method, despite its scaling and bandwidth drawbacks. MVPN-driven EVPN multicast shines mainly in large multi-pod data centres where the sheer volume of multicast traffic justifies the extra engineering effort. As ASICs become increasingly capable and vendor code more closely tracks published drafts, the combined approach is likely to become both practical and economically attractive (1).

8. Case Study: Equity Trading Colocation Deployment

8.1 Overview of the Environment

Inside a busy colocation tower like Equinix NY4 in Northern New Jersey, a top-tier equity-trading firm began a large-scale upgrade of its technical infrastructure. The building houses both exchange matching engines and client servers, meaning that even small gains in signal propagation time can translate into a valuable competitive edge. To capitalise on this advantage, the firm had arranged purpose-built bare-metal trading boxes, market-data gates, and risk-analysis engines in a tightly packed, low-latency layout. The original scheme relied on direct Layer 2 links between key devices and heavily utilized classic multicast to distribute market data to clients (31).

8.2 Migration Path from MPLS L2VPN to VXLAN EVPN

The older network backbone relied on MPLS Layer 2 VPNs to stretch the same broadcast domain across several data halls. That approach worked well in smaller setups, yet rapidly became unwieldy once multicast groups multiplied and traffic soared. Operators soon noticed scaling limits, cumbersome configuration procedures, and weak tenant isolation, all of which pushed the team to explore a contemporary VXLAN EVPN design. Migration commenced with a carefully phased deployment in which VXLAN tunnels were provisioned between racks, each relying on dual-homed VTEPs mounted on the top-of-rack switches. To streamline control-plane operations, route reflectors were sited centrally within the core, thereby facilitating a single domain for EVPN route distribution. The arrangement enabled engineers to assign distinct VNIs for segmentation, to advertise host and multicast memberships dynamically, and ultimately to lay the groundwork for future multi-tenant scenarios.

8.3 Leaf-Spine Topology, Dual-Homed VTEPs, Route Reflectors

The revised fabric adhered to a leaf-spine architecture, delivering predictable latency alongside multiple redundant forwarding paths. Every leaf switch functioned as a VTEP and connected to two spine switches, guaranteeing elevated availability. These VTEPs used EVPN Route Types 2 and 6 to announce host addresses and multicast-group identifiers. Route reflectors propagated policies consistently, and control-plane convergence was scrutinized during market simulations to confirm failover performance. The new design permitted Layer 2 extension atop a routed underlay while also leveraging BGP-driven learning, distributed ARP suppression, and multicast signaling that scales with the number of peers. Initial trials at light traffic levels yielded encouraging throughput, yet the system soon faced testing under full-market load.

As shown in the figure below, the spine-leaf layout—combined with dual-homed VTEPs and route reflectors—formed a scalable and resilient backbone for multicast-aware EVPN overlays in trading colocation networks.

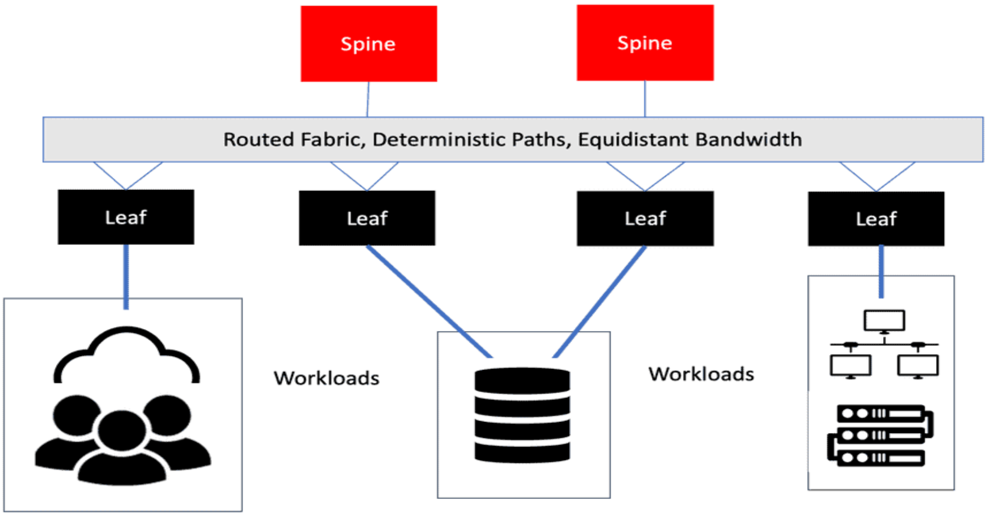

***Figure 6**: spine-leaf-architecture*

8.4 Observed Multicast Issues: Jitter, Replication Delays, VTEP Overload

Immediately following deployment, multicast traffic replicated through Ingress Replication revealed operational limits (5). Performance degraded most dramatically during the market peak, centered on the New York Stock Exchange open at 09:30 and close at 16:00, when market-data bursts soared past ten thousand packets per second. VTEPs tasked with multicast initiation entered CPU saturation, memory buffers filled, and subsequent packet drops occurred (19). Telemetry indicated increased jitter, disordered packet streams, and intermittent feed outages. Market-data handlers recorded time-stamp divergences between redundant receivers, rendering reference clocks unreliable for algorithmic signal generation. Software-only switches—heavily burdened by the replication workload—sustained the worst impact, failing to deliver packets at consistent rates and introducing noticeable latency variation.

8.5 Post-Mitigation Results and Tuning Strategies

Restoring low-latency multicast required a multi-faceted optimization. First, the firm assigned high-rate feeds to dedicated VNIs, confining each replication domain to its primary consumers and limiting extraneous duplication. Second, hardware-accelerated switches assumed primary responsibility; upgraded firmware re-targeted replication logic to ASIC pipelines, freeing system CPU for forwarding tasks. As a final safeguard, telemetry thresholds and alert rules were tightened, enabling real-time kernel-based corrections during future bursts. Preliminary analysis indicates that the median jitter has been reduced to below five milliseconds, and packet loss is approaching zero.

Selective BUM replication was activated, allowing each VTEP to forward only those frames whose ingress port either announced membership via IGMP or carried an EVPN Route Type-6. Quality-of-service profiles were updated in real-time with live telemetry, pinpointing congestion wherever it occurred and applying low-latency queuing when necessary. At the same time, a cautious pilot migrated low-risk feeds to native multicast routed over PIM-SM within the underlay. Early measurements indicated lighter replication load and improved bandwidth use, thanks to quicker pruning and shorter active distribution trees. Subsequent observation confirmed a stable operating window. Latency spikes fell well within microsecond targets, jitter approached zero, and the overall replication load spread evenly across the fabric. As a result, multicast packets again met the timing thresholds required by the firm's automated trading platforms. The exercise, therefore, illustrates that although VXLAN-EVPN multicast can strain financial networks, disciplined design and hardware-specific tuning can reliably restore performance to the levels expected in high-frequency environments (14).

9. Methodology: How Challenges Were Evaluated

Investigating the multicast scaling constraints present in VXLAN/BGP EVPN deployments, particularly within latency-sensitive trading colocation facilities, demanded an orderly, data-driven examination. Laboratory simulations, fine-grained packet analyses, and interviews with network architects who regularly manage these infrastructures together established a robust framework for isolating bottlenecks and testing candidate remedies.

9.1 Lab Setup with Cisco, Arista, and Juniper Hardware

The experimental arena mirrored commercial operations by combining branded hardware—

Cisco's Nexus series, Arista's 7280 and 7050X models, and Juniper QFX switches—with suitably licensed software alternatives. Where direct access to racks proved impractical, vendor virtual network operating systems, such as Nexus 9000v, Arista vEOS, and vQFX, ran inside hypervisor-managed machines that replicated core data-center characteristics. Adopting a leaf-spine architecture, the topology tracked the standard blueprints found in trading houses. Bridging endpoint VLAN tunnel endpoints (VTEPs), spine nodes, and control-plane route reflectors distribute multicast membership across several overlay network identifiers (VNIs). Linux hosts and dedicated test instruments mimicked client subscriptions, allowing for the consistent measurement of replication domain size, packet arrival rate, and underlay transparency while comparing throughput, latency, and failure-resilience metrics across the selected platforms.

9.2 Multicast Traffic Simulation: Tools Used

To reproduce realistic trading behavior, software capable of sustaining high-rate multicast streams over extended periods was required. For the bulk of the experiments, two industry-grade appliances—Ixia Breaking Point and Spirent Test Center—were employed. These platforms generate thousands of multicast packets per second across hundreds of groups and offer precise control over packet size, inter-arrival jitter, burst length, and other parameters. This made it possible to mimic busy market openings or sudden spikes during periods of heightened volatility (17).

Where custom behavior was necessary, Scapy served as an agile, script-driven tool. Tests were written to simulate IGMP join-and-leave sequences, inject malformed headers, and probe the resilience of each VTEP's multicast processing stack. To mirror a realistic subscriber environment, Linux containers and virtual machines received the packets, forming a distributed ring of consumers layered over the VXLAN tunnel. This blend of commercial testing platforms and open-source scripting provided both repeatable benchmarks and flexible ad-hoc validation.

9.3 Measured Metrics: Latency, Jitter, CPU Demand, and Failover Response

Testing emphasized metrics essential to low-latency trading operations. Latency was recorded from feed source to listener by timestamping at microsecond granularity on both ends. Jitter was derived from the variance in packet-inter-arrival times, revealing swings in replication cadence. CPU-utilization data stems directly from the VTEPs, especially those processing incoming replicas; these readings determine the extra load HER adds as membership grows. Failover was assessed by forcing link or node outages and timing the takeover of backup VTEPs. In PIM-SM topologies, special attention was given to the interval required for the Rendezvous Point to restart forwarding and for clients to transition from shared to source-specific trees. All measurements were repeated under light, moderate, and heavy traffic to show how the loading colors recovery speed.

9.4 Packet-Level Analysis: Wireshark, tcpdump, IGMP Joins and Leaves

For fine-grained visibility into multicast operations, bright-colored analysis tools such as Wireshark and tcpdump were placed at key points across the switch fabric. Captured frames confirmed the expected tunnels, revealed unexpected replication loops, and recorded the signaling flood when hosts join or leave a group. Particular attention was directed toward IGMP join and leave messages initiated by simulated endpoints (11). These control packets enabled the team to verify whether each VTEP refreshed its multicast distribution table as expected and to determine if any extraneous replication reached the data plane. Packet loss and duplication were subsequently recognized by analyzing sequence numbers on UDP-based multicast streams. By decoding outer VXLAN headers and mapping the multicast payload spread, researchers reverse-engineered the ingress replication paths.

9.5 Insights from Expert Interviews and Vendor Documentation

To supplement laboratory measurements, structured interviews were conducted with network architects and engineers employed by large financial firms and relevant hardware vendors. Participants described operational constraints—regulatory obligations, production load-balancing practices, and platform-specific quirks—that lab environments cannot fully reproduce. Drifting in parallel, deployment guides and technical documentation from Cisco, Arista, and Juniper were reviewed to enumerate recommended configurations, hardware limits, and planned features. Drafts from the IETF BESS working group—particularly RFC 7432 and the draft titled *ietf-bess-evpn-igmp-mld-proxy*—have been instrumental in clarifying EVPN multicast signaling mechanics and outlining the evolution of relevant standards. Taken together, this multi-pronged approach yielded a robust picture of multicast behavior in VXLAN EVPN fabrics and confirmed the performance characteristics seen in live networks.

As illustrated in the figure below, the methodology combined technical evaluation, field experience, and standards alignment to produce a well-rounded performance profile.

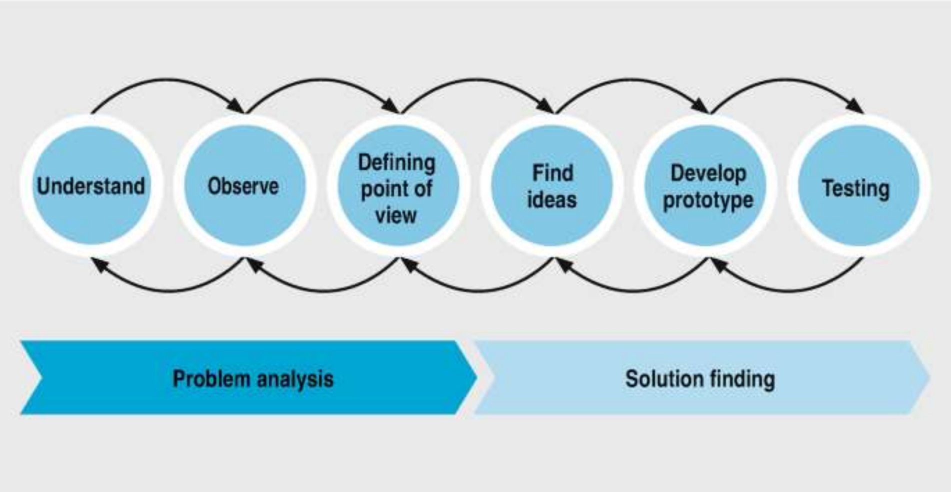

***Figure 7**: Methodology*

10. Optimization Strategies and Best Practices

Multicast replication and its impact on scalability remain pressing concerns in VXLAN/BGP EVPN trading networks, primarily where strict requirements for data integrity and near-zero latency exist (34). While these demands can strain fabric resources, thoughtful design and hardware-aware deployment steps substantially alleviate the burden. The following paragraphs summarize optimization methods validated in both production exchanges and advanced test labs, all aimed at sustaining multicast throughput as the node count grows.

10.1 Smart Replication: Selective BUM Handling per VNI

Delivering broadcast, unknown unicast, and multicast (BUM) frames indiscriminately to every VTEP creates predictable bottlenecks. The refinement that addresses this problem is selective replication, in which each multicast stream is forwarded only to those devices that have expressed interest through per-VNI control signalling. Instead of flooding the entire VXLAN segment, the fabric consults Inclusive Multicast Ethernet Tag (IMET) routes, builds targeted replication trees, and uses the EVPN control plane to update these lists whenever group membership changes.

This method reduces unnecessary overhead by minimizing traffic duplication. Suppose a virtual network identifier supports three multicast groups, yet only two virtual tunnel endpoints need the same stream. In that scenario, the replication list registers only those two VTEPs. By doing so, the source VTEP spends fewer CPU cycles, and the wider spine-leaf fabric uses bandwidth more judiciously. Adding careful Internet Group Management Protocol (IGMP) snooping at the edge reinforces this efficiency; it ensures that control-plane adjustments respond solely to active listeners.

10.2 Hardware Acceleration: Offload HER to ASICs

When head-end replication proves unavoidable—especially in settings where standard multicast cannot be turned on—pushing the work into hardware assets becomes indispensable. Contemporary switches built around application-specific integrated circuits, whether Broadcom Trident 3, Trident 4, or Jericho 2, can execute HER directly in silicon. That capability boosts throughput markedly while easing the burden on the switches' general-purpose processor. Such ASIC-based VTEPs can duplicate packets at line rate, a non-negotiable feature in trading floors, where bursts sometimes exceed 5,000 packets per second for a single group. Still, accurate configuration and subsequent checks are vital because replication efficiency varies among switch models. Administrators should therefore examine vendor documentation and capability matrices in detail, aligning hardware limits with the expected traffic profile.

10.3 Native Multicast: PIM-SM with IGMP Snooping and Anycast RP

When infrastructure resources are available, implementing native multicast through Protocol Independent Multicast – Sparse Mode (PIM-SM) in the underlay generally scales better than hierarchical Ethernet Relay. This architecture delegates packet replication to the network by constructing multicast distribution trees, hence relieving source VTEPs of heavy forwarding duties. By enabling IGMP snooping on VTEPs, the fabric admits only those ports that have active subscribers into the tree, achieving targeted rather than blanket replication (32). Employing Anycast Rendezvous Points further sharpens convergence periods and fault tolerance, since multiple PIM-SM RPs can serve requests while the network sees a unified logical address. The trade-off is added complexity in the underlay control plane. Still, the payoff is substantial: link utilization is more predictable, and the chance of uncontrolled multicast storms is markedly reduced. In use cases with stable membership and high event rates, such as low-latency financial feeds, the combined technique delivers reliable and repeatable bandwidth growth.

10.4 Segmentation Techniques: Per-Tenant Isolation and Hierarchical Overlays

Within multi-tenant collocated trading centers, segmenting multicast domains by tenant or service preserves operational control and guarantees equitable bandwidth for all users. Allocating a separate virtual network identifier (VNI) to each trading team eliminates overlapping group addresses and enables consistent policy enforcement across both the control and data planes. To accommodate growth in participant numbers, hierarchical overlays can be layered on top of this scheme. A global base VNI distributes market data to every tenant, while dedicated overlay VNIs transport proprietary analytics or execution signals. This setup maintains clear logical boundaries while allowing targeted information sharing when necessary. Route-target filtering on control-plane advertisements further curbs the risk of unwanted routing leaks, easing both security audits and routine diagnostics. Associating each tenant with its own Virtual Routing and Forwarding (VRF) instance deepens isolation. It simplifies compliance verification, a crucial feature in regulated environments where every access boundary must be thoroughly documented.

As shown in the figure below, this multi-tiered segmentation model ensures that each tenant operates within its own secure and logically distinct enclave, even while sharing underlying physical infrastructure

.

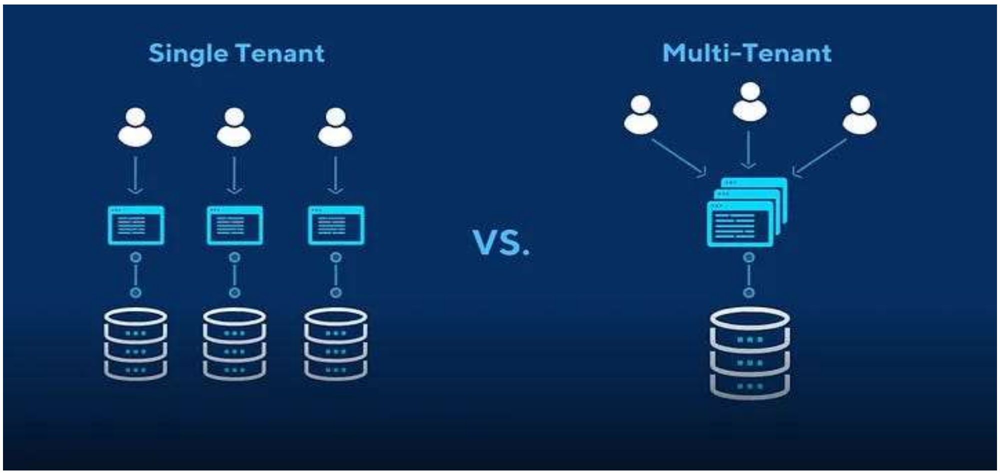

Figure 8: *Multi-Tenant Architecture*

10.5 Real-Time Monitoring: sFlow, Telemetry Pipelines, AI/ML Forecasting

Robust multicast optimization requires a visibility loop, not a static dial. sFlow, NetFlow, and ERSPAN-based telemetry pipelines generate per-second snapshots of replication behaviour. From those snapshots, operators assign packet counts to individual multicast groups, chart replication paths across VTEPs, and extract interface-level utilization. When paired with InfluxDB, Grafana, and Telegraf, the resulting data pipeline transforms raw numbers into sharp, time-series graphs, alerting engineers to outliers before they develop into outages. Some production environments have begun testing machine-learning models that forecast multicast feed pressure by correlating incoming pulse streams with historical workloads. For example, if the market-open burst of TOPS usually overloads VTEP-3, a trained classifier can flag that node minutes in advance and trigger either a route preemption or an upstream policing policy, thereby keeping latency within SLAs. Telemetry also assists fault identification. Drastic drops in join/leave count, combined with skewed replication ratios, frequently signal silent failures within the tree. Real-time hooks wired to incident-management platforms then surface the problems as tickets, not as post-mortems, preserving revenue and trader confidence.

11. Security and Compliance in Multicast Trading

In fast-moving trading firms, where every millisecond counts and sensitive data must remain secure and intact, safe multicast delivery is just as crucial as quick and reliable delivery (3). Allowing multicast packets to slip through without proper security measures opens the door to honest mistakes and malicious actors eager to intercept or tamper with the stream.

11.1 Threat Models: Group Leakage, Unauthorized Joins, Broadcast Storms

Because multicast sends the same bit of information to multiple receivers simultaneously, it poses unique risks to a high-pressure trading backbone. Group leakage occurs when a multicast message intended for one team is accidentally sent to an unrelated server, typically due to incomplete address lists or loose replication settings. Unauthorized joins, whether invited by a careless administrator or surreptitiously inserted by a clever attacker, can expose sensitive data or disrupt orderly systems. Finally, although rare in modern EVPN setups, broadcast storms may still occur if permission rules are too lax or IGMP snooping fails, resulting in unnecessary traffic flooding every VTEP and slowing network speed.

As illustrated in the figure below, each threat introduces a distinct failure path, and mitigating these risks requires robust network segmentation, access control lists (ACLs), and consistent monitoring of multicast group memberships.

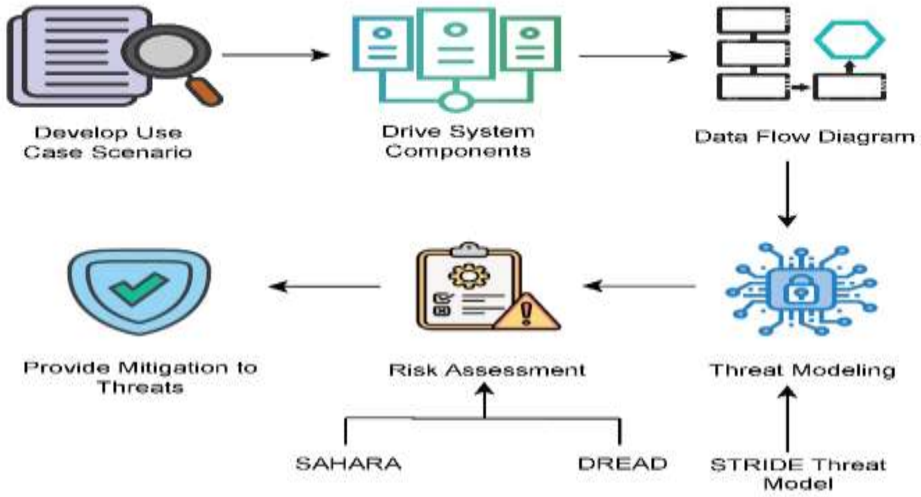

***Figure 9**: The step-by-step research methodology.*

11.2 Access Control: Group ACLs, RP Filtering, IGMP Policing

Networks face serious risks from misbehaving multicast traffic, and to guard against them, engineers rely on a mix of access-control tools. Group-based Access Control Lists (ACLs) are programmed into both VTEPs and core routers, letting the operator explicitly decide who can join which multicast group based on source address, destination address, or even the tenant that owns the packets. In networks that still use native multicast and PIM-SM, Rendezvous Point (RP) filtering is added so unwanted sources cant sneak onto the tree by pretending to register through the RP. Finally, at the edge ports, IGMP policing limits each endpoint to its approved groups, keeping dynamic-join storms in check and sparing switches from running out of resources due to the simultaneous subscription to too many streams.

11.3 Tenant and VNI Isolation: VRF per Customer, Route-Target Filtering

Real multicast segmentation relies on dividing tenants by their Virtual Network Identifiers (VNIs) and maintaining completely separate Virtual Routing and Forwarding (VRF) tables. When each trading firm or app receives its own VRF, all control-plane routes—including EVPN Route Type 6 multicast advertisements—remain isolated within that box and never bleed into someone else's domain. To clean the exits, route-target filtering running on route reflectors allows only designated updates to cross the border, blocking careless imports or exports that would otherwise misroute traffic and compromise the strict isolation that these sensitive environments demand (4). Keeping different business lines on separate virtual roads does more than boost security; it also respects company secrets and regulatory rules. When several banks share the same physical wires, slicing the control and data lanes helps protect competition and ensures everyday operations run smoothly.

11.4 Compliance Mandates: MiFID II, Audit Trails, Deterministic Paths

Regulations like MiFID II require firms to record every trade message, stamp the exact arrival time, and deliver it within a set tick. As soon as multicast gets invited to the party, the underlying network must meet those speed limits while still logging who-said-what when. Each time a market data feed pings out, every receiver should receive the same message in the same order; if a packet is missed or shuffled, and the audit clock starts ticking. Hitting these targets means the team has tight control over how packets are copied and routed through the switch. Network managers must be ready to show a clear trail that proves each packet hopped the right way, spell out the rules behind each move, and explain how backup plans kick in-all without letting random delays sneak in. Since regulators now follow the data line from start to finish, logs, telemetry, and rule checks have gone from nice-to-have tools to must-show proof.

11.5 Logging and Forensics: Traffic Replay, Source Attribution

When something goes wrong—whether it's slow performance, lost data, or a rule that can't be proven—it's vital to replay the network traffic and identify where the issue originated. To achieve this, teams utilize innovative logging tools that monitor multicast flows, track every group join and leave, and record how VTEPs replicate packets, all while the event is occurring. In some places, these multicast streams are sent directly to forensic storage, allowing every single packet to be examined later if regulators or courts require proof. Pinning down the source is equally important. In a busy shared multicast network, knowing which VTEP or application initiated a data stream and how far it traveled helps resolve problems and meet compliance requirements. Engineers get that deep insight by combining EVPN route checks, IGMP traceback records, and telemetry dumps that feed into security information and event management, or SIEM, consoles.

12. Future Outlook and Recommendations

Trading networks are growing faster than ever, with mountains of new data and tight deadlines pushing every link in the chain (6). To keep up, multicast delivery systems must also step up their efforts. VXLAN, paired with BGP-eVPN, already does a solid job of extending Layer 2 reach with a smart control plane; however, real scalability still hinges on how multicast streams behave in both the overlay and the underlying world, requiring fresh technologies and design. Strategies are now transforming the multicast landscape within financial trading centers.

12.1 Emerging Technologies: Programmability and Path Control

A handful of new ideas promise to enhance multicast performance and bolster network stability as the load increases. First on the list is Segment Routing over IPv6 (SRv6), which provides engineers with a flexible and programmable way to steer traffic. By building defined, nimble paths instead of the usual multicast trees, SRv6 reduces failover lag and prevents networks from relying on fixed rendezvous points that can quickly become bottlenecks. P4-programmable switches add another layer of promise. Because these boxes enable operators to write packet-handling rules that run within the data plane, they can establish group logic, drop duplicates, and reroute flows based on real-time network conditions without waiting for the control plane to catch up. That kind of freedom opens doors to congestion-aware copying, time-window priorities, and many other tricks once thought too costly or slow to use in production.

12.2 Application-Layer Alternatives: Kafka and NATS

Although multicast sits squarely at the network layer, Apache Kafka and NATS are now widely used in hybrid and cloud settings to distribute data feeds (16). Both systems promise dependable message delivery, support event replay, and integrate seamlessly with distributed computing pipelines. They do add some processing cost and still lag behind bare-metal multicast in ultra-tight latency cases; yet, they shine in scenarios such as market-data replay, back-testing trading strategies, and any analysis that prioritizes correctness over speed. Additionally, by routing packets over unicast tunnels, they avoid the multicast headaches that often hinder virtual machines and containers.

12.3 Recommendations for Network Designers

Network engineers who want dependable, scalable multicast in an EVPN fabric should keep the following tips in mind: Design for Isolation and Containment. Assign separate Virtual Network Identifiers and Virtual Routing and Forwarding tables to each tenant or service, ensuring that multicast traffic for one group does not overlap with that of another. This reduces unnecessary packet copies and protects sensitive streams. Minimize HER Scope. Hop-by-hop Replication may work fine for low-rate groups, but high-throughput feeds will grind it to a crawl. Whenever the network can handle it, switch back to native multicast in the underlay, allowing the hardware to do the heavy lifting. Set Up Group-Based QoS and Telemetry: First, sort multicast streams into groups and apply the proper QoS rules for each set, depending on the importance of the content. Then, monitor group health in real-time using tools such as sFlow, IPFIX, or gRPC streaming to immediately identify congestion points or anomalous behavior.

Test before You Deploy: Before going live, fire up traffic testers such as Ixia, Spirent, or even Scapy to mimic streams at their heaviest expected rates. Verify that every component of the path, including VTEPs, switches, and routers, maintains packet flow with minimal drops and jitter.

Leverage Hardware Offload: Whenever possible, choose switches with built-in multicast offload capabilities, such as the Broadcom Trident or Jericho range. Offloading those tasks to hardware takes pressure off the software plane and helps handle sharp traffic bursts cleanly.

Plan for Failover: Failover and Convergence: Map the multicast network with failure zones in mind, and utilize Anycast RP, BFD on PIM links, and quick timer settings to ensure reroutes kick in almost instantly when a link or node fails.

12.4 Evolving but Enduring: The Role of Multicast

Even with pub/sub systems and flashy programmable networks winning headlines, multicast still rules the market-data beat. Its knack for pushing fresh prices to hundreds of traders in one sweep, without clogging the network, beats most alternatives in terms of bandwidth and lag. Instead of being pushed aside, multicast is getting smart upgrades that patch the leg-cy holes folks used to complain about. In the lightning-fast world of trading, a multicast setup requires precise blueprints, sink-or-swim simulations, and continuous real-time checks once live (24). Marrying tried-and-true networking rules with new-school programmability and observability keeps the entire system agile yet rock-solid.

13. Conclusion

The architecture guiding contemporary trading infrastructures increasingly relies upon VXLAN paired with BGP EVPN, setting the benchmark for scalable, policy-aware networks in financial colocation facilities. Such settings demand timing accuracy, persistent availability, and responses measurable in microseconds, attributes that expose the weaknesses of conventional flat Layer-2 fabrics in both scale and security. By constructing Layer-2 overlays over a routed Layer-3 spine, VXLAN and BGP EVPN facilitate dynamic MAC and IP learning, allowing for fine-grained tenant isolation. These features provide trading venues with the controlled, discrete, and adaptable conduits necessary for low-latency algorithms. Within these latency-sensitive environments, multicast emerges as a critical service mechanism. It supplies the medium through which high-frequency market broadcasts—equity prices, option ticks, live order books, and supporting reference data—traverse the network. Because multicast transmits a single stream that multiple endpoints can tap concurrently, it economizes on bandwidth, reduces processor load, and shortens transit time relative to distinct unicast feeds to each receiver. For strategies reliant on instantaneous and uniform data arrival, multicast delivers consistency, accelerates decision cycles, and imposes minimal additional delivery overhead.

Although multicast offers clear benefits in large-scale forwarding, bringing it into contemporary VXLAN EVPN fabrics creates significant operational friction. The first question that arises on any deployment is whether to rely on ingress replication (HER) or to implement a native multicast architecture within the underlay. With HER, every multicast frame is duplicated at the ingress VTEP and sent to each interested receiver in unicast streams. This approach demands little from the underlay, permits rapid execution, and eventually becomes the path of least resistance for many operators. Yet, empirical performance tests reveal a troubling breaking point: during high-volume conditions—such as early market openings or major macroeconomic releases—the sheer number of streams can overload the originating VTEP, spike bandwidth consumption on uplink pairs, and introduce jitter or dropped packets that traders cannot afford to tolerate.

Native multicast leverages PIM, MVPN, or related protocols to build shared trees that limit replication to upstream points on the network. Cumulatively, these techniques minimize link occupancy, reduce CPU load on source VTEPs, and scale more gracefully as subscriber numbers grow. That scalability, however, is often offset by a lack of design discipline. Architects must address PIM convergence timing, carefully position rendezvous points, enforce inter-VRF policies, and ensure transparent failover —competencies that often require additional configuration lines and continuous verification burdens. Early-stage pilots report benefits, yet ongoing support teams tend to encounter the very complexity that was promised to be left behind. Striking an effective balance among simplicity, scalability, and performance is indispensable in any network design. In VXLAN-EVPN environments, optimal multicast delivery emerges only through thoughtful architectural choices that are informed by empirical performance data and tailored to the precise requirements of financial trading systems.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Alcaín, E., Fernandez, P. R., Nieto, R., Montemayor, A. S., Vilas, J., Galiana-Bordera, A., ... & Torrado-Carvajal, A. (2021). Hardware architectures for real-time medical imaging. Electronics, 10(24), 3118. https://doi.org/10.3390/electronics10243118

[2] Alshaer, H. (2015). An overview of network virtualization and cloud network as a service. International Journal of Network Management, 25(1), 1-30. https://doi.org/10.1002/nem.1882

[3] Alshammari, A. R. (2020). Resilient Wireless Network Virtualization with Edge Computing and Cyber Deception (Doctoral dissertation, Howard University).

[4] Balbaa, M. E. (2022). International Transport Corridors. Tashkent State University of Economics: Tashkent, Uzbekistan.

[5] Bhardwaj, K., & Nowick, S. M. (2018). A continuous-time replication strategy for efficient multicast in asynchronous NoCs. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 27(2), 350-363. https://doi.org/10.1109/TVLSI.2018.2876856

[6] Blanchard, D. (2021). Supply chain management best practices. John Wiley & Sons.

[7] Brogaard, J., Hagströmer, B., Nordén, L., & Riordan, R. (2015). Trading fast and slow: Colocation and liquidity. The Review of Financial Studies, 28(12), 3407-3443. https://doi.org/10.1093/rfs/hhv045

[8] Cannarella, A. (2022). Multi-Tenant federated approach to resources brokering between Kubernetes clusters (Doctoral dissertation, Politecnico di Torino). http://webthesis.biblio.polito.it/id/eprint/25422

[9] Chavan, A. (2021). Exploring event-driven architecture in microservices: Patterns, pitfalls, and best practices. International Journal of Software and Research Analysis. https://ijsra.net/content/exploring-event-driven-architecture-microservices-patterns-pitfalls-and-best-practices

[10] Chavan, A. (2022). Importance of identifying and establishing context boundaries while migrating from monolith to microservices. Journal of Engineering and Applied Sciences Technology, 4, E168. http://doi.org/10.47363/JEAST/2022(4)E168

[11] Chen, L. (2017). Performance Evaluation for Secure Internet Group Management Protocol and Group Security Association Management Protocol (Doctoral dissertation, Concordia University). https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?idNumber=1135022369

[12] Dhanagari, M. R. (2024). MongoDB and data consistency: Bridging the gap between performance and reliability. Journal of Computer Science and Technology Studies, 6(2), 183-198. https://doi.org/10.32996/jcsts.2024.6.2.21

[13] Dhanagari, M. R. (2024). Scaling with MongoDB: Solutions for handling big data in real-time. Journal of Computer Science and Technology Studies, 6(5), 246-264. https://doi.org/10.32996/jcsts.2024.6.5.20

[14] Emami, M., Bayat, A., Tafazolli, R., & Quddus, A. (2024). A survey on haptics: Communication, sensing and feedback. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2024.3444051

[15] Fawcett, R. L. (2024). The Contours of the Cloud: Dissecting the Real Estate Investment Decisions of Data Center Operators (Doctoral dissertation, Massachusetts Institute of Technology). https://hdl.handle.net/1721.1/157114

[16] George, J. (2022). Optimizing hybrid and multi-cloud architectures for real-time data streaming and analytics: Strategies for scalability and integration. World Journal of Advanced Engineering Technology and Sciences, 7(1), 10-30574. https://ssrn.com/abstract=4963389

[17] Ghosh, S. (2023). Building Low Latency Applications with C++: Develop a complete low latency trading ecosystem from scratch using modern C++. Packt Publishing Ltd.

[18] Goel, G., & Bhramhabhatt, R. (2024). Dual sourcing strategies. International Journal of Science and Research Archive, 13(2), 2155. https://doi.org/10.30574/ijsra.2024.13.2.2155

[19] Karwa, K. (2023). AI-powered career coaching: Evaluating feedback tools for design students. Indian Journal of Economics & Business. https://www.ashwinanokha.com/ijeb-v22-4-2023.php

[20] Kodheli, O., Lagunas, E., Maturo, N., Sharma, S. K., Shankar, B., Montoya, J. F. M., ... & Goussetis, G. (2020). Satellite communications in the new space era: A survey and future challenges. IEEE Communications Surveys & Tutorials, 23(1), 70-109. https://doi.org/10.1109/COMST.2020.3028247

[21] Konneru, N. M. K. (2021). Integrating security into CI/CD pipelines: A DevSecOps approach with SAST, DAST, and SCA tools. International Journal of Science and Research Archive. Retrieved from https://ijsra.net/content/role-notification-scheduling-improving-patient

[22] Kumar, A. (2019). The convergence of predictive analytics in driving business intelligence and enhancing DevOps efficiency. International Journal of Computational Engineering and Management, 6(6), 118-142. Retrieved from https://ijcem.in/wp-content/uploads/THE-CONVERGENCE-OF-PREDICTIVE-ANALYTICS-IN-DRIVING-BUSINESS-INTELLIGENCE-AND-ENHANCING-DEVOPS-EFFICIENCY.pdf

[23] Mirtl, M., Borer, E. T., Djukic, I., Forsius, M., Haubold, H., Hugo, W., ... & Haase, P. (2018).

Genesis, goals and achievements of long-term ecological research at the global scale: a critical review of ILTER and future directions. Science of the total Environment, 626, 1439-1462. https://doi.org/10.1016/j.scitotenv.2017.12.001

[24] Morel, L. P. (2017). Using ontologies to detect anomalies in the sky. Ecole Polytechnique, Montreal (Canada). https://www.proquest.com/openview/1310c97e55ee11adc005c478ad646164/1?pq-origsite=gscholar&cbl=18750

[25] Nyati, S. (2018). Revolutionizing LTL carrier operations: A comprehensive analysis of an algorithm-driven pickup and delivery dispatching solution. International Journal of Science and Research (IJSR), 7(2), 1659-1666. Retrieved from https://www.ijsr.net/getabstract.php?paperid=SR24203183637

[26] Prasad, P., Mohammad, T., & Sainio, P. (2024). Enhancing Security in Software-Defined Networking (SDN) based IP Multicast Systems: Challenges and Opportunities. https://www.utupub.fi/bitstream/handle/10024/178222/Prasad_Preety_Masters_Thesis.pdf?sequence=1

[27] Raju, R. K. (2017). Dynamic memory inference network for natural language inference. International Journal of Science and Research (IJSR), 6(2). https://www.ijsr.net/archive/v6i2/SR24926091431.pdf

[28] Sardana, J. (2022). The role of notification scheduling in improving patient outcomes. International Journal of Science and Research Archive. Retrieved from https://ijsra.net/content/role-notification-scheduling-improving-patient

[29] Singh, V. (2023). Federated learning for privacy-preserving medical data analysis: Applying federated learning to analyze sensitive health data without compromising patient privacy. International Journal of Advanced Engineering and Technology, 5(S4). https://romanpub.com/resources/Vol%205%20%2C%20No%20S4%20-%2026.pdf

[30] Singh, V., Oza, M., Vaghela, H., & Kanani, P. (2019, March). Auto-encoding progressive generative adversarial networks for 3D multi object scenes. In 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (pp. 481-485). IEEE. https://arxiv.org/pdf/1903.03477

[31] Sukhadiya, J., Pandya, H., & Singh, V. (2018). Comparison of Image Captioning Methods. INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH, 6(4), 43-48. https://rjwave.org/ijedr/papers/IJEDR1804011.pdf

[32] Tafreshi, V. H. F. (2015). Secure and robust packet forwarding for next generation IP networks. University of Surrey (United Kingdom).

[33] Trestioreanu, L., Shbair, W. M., de Cristo, F. S., & State, R. (2023, May). Xrp-ndn overlay: Improving the communication efficiency of consensus-validation based blockchains with an ndn overlay. In NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium (pp. 1-5). IEEE. https://doi.org/10.1109/NOMS56928.2023.10154402

[34] Zhang, Y., Kutscher, D., & Cui, Y. (2024). Networked metaverse systems: Foundations, gaps, research directions. IEEE Open Journal of the Communications Society. https://doi.org/10.1109/OJCOMS.2024.3426098

[35] Zheng, K., Zheng, Q., Chatzimisios, P., Xiang, W., & Zhou, Y. (2015). Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions. IEEE communications surveys & tutorials, 17(4), 2377-2396. https://doi.org/10.1109/COMST.2015.2440103