American Journal of Applied Science and Technology

VOLUME Vol.05 Issue09 2025

PAGE NO. 74-87

Enhanced Multicast Scaling in Low-Latency Trading Colocation Environments: A Critical Analysis of VXLAN/BGP EVPN Architectures

Elias J. Sterling

Department of Advanced Networking and Telecommunications, Global School of Financial Engineering, London, United Kingdom

Sarah N. Reynolds

Faculty of Data Center Architecture, Institute for Computational and Financial Sciences, Singapore, Singapore

Received: 25 August 2025; **Accepted:** 15 September 2025; **Published:** 30 September 2025

Abstract

Context: High-Frequency Trading (HFT) relies critically on ultra-low latency dissemination of market data via IP Multicast within colocation facilities. The adoption of VXLAN/BGP EVPN has provided scalable Layer 2/3 virtualization for these multi-tenant environments. However, the sheer volume of market data feeds presents a significant and often overlooked challenge to the multicast scaling capabilities of standard EVPN architectures.

Objective: This paper provides a critical analysis of the scaling limitations of multicast forwarding mechanisms within VXLAN/BGP EVPN overlays, specifically examining their suitability for the stringent latency and group count demands of HFT colocation networks.

Methods: We analytically model and evaluate two primary EVPN multicast forwarding strategies: Ingress Replication (IR) and PIM-integrated Underlay Multicast. Key performance metrics, including Control-Plane Convergence Time, Data-Plane Latency Jitter, and Multicast Group Capacity (MGC), are defined and used for a comparative assessment based on typical HFT traffic profiles.

Results: Our analysis reveals that standard IR suffers from significant control-plane state proliferation (BGP EVPN Type-6 route explosion) and bandwidth inefficiency. Conversely, PIM-integrated solutions, while data-plane efficient, introduce complexity and potential for forwarding-state synchronization issues and hardware resource exhaustion (TCAM). Neither approach optimally meets the combined low-latency and high-MGC requirements.

Conclusion: The conventional implementations of VXLAN/BGP EVPN are insufficient to support the massive, low-latency multicast scaling required by modern HFT colocation. Architectural enhancements, including SDN-based control-plane optimization and innovative route aggregation techniques, are necessary to ensure the continued performance and resilience of these critical financial networks.

Keywords: VXLAN/BGP EVPN, IP Multicast Scaling, High-Frequency Trading (HFT), Colocation Networks, Low Latency, Network Virtualization, Multicast Group Capacity (MGC)

1. Introduction

1.1. Contextualizing High-Frequency Trading (HFT) and Latency Constraints

The architecture of modern financial markets is dominated by algorithmic trading, a discipline where fractions of a second can equate to significant competitive advantage. High-Frequency Trading (HFT), a specialized subset of algorithmic trading, relies fundamentally on the rapid consumption and processing of market data. For HFT firms, network latency is not merely a performance metric but a critical determinant of profitability and operational viability. The pursuit of ultra-low latency has driven a physical centralization trend, necessitating the strategic use of colocation facilities. By situating trading servers within the data centers hosting exchange matching engines, firms minimize the physical distance and, consequently, the network latency to the source of market data. This environment dictates architectural choices where every microsecond matters.

1.2. The Role of Multicast in Trading Ecosystems

The efficient distribution of real-time market data—such as price quotes, trade confirmations, and order book updates—is universally accomplished using IP Multicast. Multicast is a bandwidth-saving communication method where a source sends a single stream of data packets to a multicast address, and the network infrastructure is responsible for replicating these packets only to the interested subscribers. In a colocation environment, thousands of client applications may subscribe to market data feeds simultaneously. Without multicast, transmitting the same data stream individually (unicast) to every receiver would overwhelm both the source servers and the network fabric. The accelerating global trading volume and the proliferation of different financial instruments have led to an explosion in the number of unique multicast groups required in a single colocation, often scaling into the tens of thousands.

1.3. Evolution of Data Center Network Virtualization

Traditional colocation networks often faced significant challenges in supporting the necessary Layer 2 (L2) connectivity and multi-tenancy required by HFT clients. Extending L2 networks across large, multi-rack environments using legacy technologies (like spanning tree protocol) introduced complexity, poor scalability, and slow convergence. The need to isolate tenants while efficiently sharing a common infrastructure necessitated the adoption of network virtualization.

The VXLAN/BGP EVPN framework has emerged as the industry standard for building highly scalable, multi-tenant data center networks. VXLAN (Virtual Extensible LAN) provides the data plane encapsulation, allowing L2 frames to be tunneled over an L3 underlay using Virtual Network Identifiers (VNIs). BGP EVPN (Border Gateway Protocol Ethernet VPN) provides the control plane, dynamically distributing L2 MAC and L3 IP routing information to the VXLAN Tunnel Endpoints (VTEPs). This separation of the control plane and data plane offers the operational flexibility needed for rapid service deployment and resource isolation, making it ideal for the highly dynamic and multi-tenant nature of colocation facilities.

1.4. Statement of the Problem and Literature Gaps

While VXLAN/BGP EVPN successfully addresses general L2/L3 scaling, its suitability for the specific demands of large-scale, low-latency multicast traffic in HFT remains inadequately studied. The core issue lies in managing the sheer volume of multicast group state required by the control plane.

- Literature Gap 1: There is a lack of a comprehensive, quantitative framework analyzing the performance degradation and state exhaustion associated with VXLAN/BGP EVPN multicast forwarding mechanisms (Ingress Replication vs. PIM integration) under HFT-specific load profiles, characterized by tens

of thousands of simultaneous, transient multicast groups.

- Literature Gap 2: The existing literature does not sufficiently address the architectural trade-offs required to overcome the Multicast Group Capacity (MGC) limitations imposed by the finite Ternary Content-Addressable Memory (TCAM) available on high-speed, low-latency network hardware.
- Problem Statement: The standard VXLAN/BGP EVPN control plane mechanisms, particularly the reliance on BGP EVPN Type-6 Selective Multicast Ethernet Tag (SMET) routes, are susceptible to control-plane state bloat and slow convergence, which translates directly into unacceptable jitter and reduced operational MGC in low-latency trading colocation environments.

1.5. Research Objectives and Paper Structure

This paper aims to critically analyze the fundamental architectural scaling limits of VXLAN/BGP EVPN multicast in the context of HFT colocation and propose innovative architectural enhancements. Specifically, the objectives are: (1) to define and model the MGC constraints of standard EVPN multicast, (2) to compare the performance trade-offs of the dominant forwarding methods (IR and PIM-integrated), and (3) to propose and evaluate control-plane optimizations utilizing Software-Defined Networking (SDN) to achieve massive MGC and low jitter.

The remainder of the paper is structured as follows: Section 2 reviews the core EVPN multicast mechanisms. Section 3 presents a comparative evaluation and scaling results. Section 4 discusses the implications and proposes novel SDN-based enhancements. Section 5 concludes the findings.

2. Methods and Architectural Analysis

2.1. Foundational Review of VXLAN/BGP EVPN Components

VXLAN/BGP EVPN operates as an overlay network built atop an existing L3 underlay. The VXLAN data plane encapsulates L2 Ethernet frames within UDP packets, identifying the logical L2 network via a 24-bit VNI and forwarding the packet through the L3 underlay to the destination VTEP. The BGP EVPN control plane is responsible for distributing reachability information necessary to establish these VXLAN tunnels.

Key EVPN routes relevant to multicast include:

- Type-2 (MAC/IP Advertisement) Route: Used by VTEPs to advertise the MAC and IP addresses of connected endpoints.
- Type-3 (Inclusive Multicast Ethernet Tag - IMET) Route: Used for discovering all VTEPs belonging to a specific VNI. This route establishes the list of VTEPs that should receive Broadcast, Unknown Unicast, and unknown Multicast (BUM) traffic. It is the basis for Ingress Replication.
- Type-6 (Selective Multicast Ethernet Tag - SMET) Route: Used to signal selective multicast interest. A VTEP interested in a specific stream will advertise a Type-6 route for the VNI to which the stream belongs.

2.2. Multicast Forwarding Mechanisms in BGP EVPN

The EVPN architecture supports two primary methods for handling IP multicast traffic, each presenting distinct trade-offs in terms of control plane complexity versus data plane efficiency.

2.2.1. Ingress Replication (IR) Multicast

Ingress Replication is the default and simplest form of multicast handling in EVPN. When a source VTEP receives an L2 multicast frame:

1. The source VTEP looks up the destination VNI.
2. It uses the list of VTEPs learned via the Type-3 (IMET) or Type-6 (SMET) routes.
3. The source VTEP then creates a separate VXLAN encapsulated unicast packet for every destination VTEP in the list. This means replication occurs entirely at the source (ingress) VTEP.

Analysis: IR is easy to deploy as it requires no multicast capability in the L3 underlay. The key drawback, however, is its inherent inefficiency. The duplication of traffic at the ingress point consumes significantly more bandwidth on the spine-to-spine and spine-to-leaf links compared to native L3 multicast. More critically for HFT, the dependence on Type-6 routes for selective forwarding leads to control-plane state explosion when (group count) is high, directly impacting convergence and increasing the potential for jitter.

2.2.2. PIM-based EVPN Integration

The second approach is to integrate the EVPN overlay with a standards-based L3 multicast protocol, typically Protocol Independent Multicast (PIM), running in the L3 underlay.

1.Multicast Tunnel (MT): The Type-3 IMET route is used to signal the existence of a Multicast Tunnel. This MT can be a PIM Shared Tree (PIM-SM) or Bi-directional PIM (BiDir-PIM) tunnel.

2.Encapsulation: The source VTEP encapsulates the multicast traffic into a VXLAN packet, which is then further encapsulated by the underlay's L3 multicast protocol (PIM).

3.Underlay Replication: The L3 underlay handles the efficient, one-to-many replication of the packet, minimizing bandwidth usage compared to IR.

Analysis: This method offers far greater data-plane efficiency. However, it introduces significant control-plane complexity. The VTEPs must now manage both the EVPN state and the PIM state, and the network must handle the complex mapping between the EVPN L2 VNI multicast groups and the L3 PIM underlay groups. Furthermore, state synchronization issues between the overlay and underlay can lead to brief periods of packet loss or Head-of-Line Blocking (HOLB) on the VTEPs, which are unacceptable in a low-latency environment.

2.3. Characterization of HFT Colocation Traffic Patterns

The HFT traffic profile imposes extreme demands that exacerbate the challenges in EVPN.

- High-Volume, Bursty Traffic: Market data feeds are highly granular, resulting in bursty packet flows (e.g., during market openings or high volatility events).
- One-to-Many Distribution: A single exchange feed is consumed by hundreds of processes, leading to the massive fan-out requirement characteristic of multicast.
- Extreme Latency Sensitivity: Packet latencies must be consistently below 50 microseconds, and Latency Jitter—the variation in latency—must be minimized, as unpredictable delays can trigger risk management alerts or lead to missed trading opportunities.
- High Group Density: Due to the variety of markets, asset classes, and data granularity, a single colocation fabric may need to support active multicast streams.

2.4. Proposed Analytical Framework: Performance Metrics

To compare the suitability of the architectures, we define three critical performance metrics:

1.Multicast Group Capacity (MGC): The maximum number of unique multicast streams () a VTEP or the entire fabric can simultaneously support while maintaining a stable control plane. This is directly limited by hardware TCAM capacity.

2.Data-Plane Latency Jitter (): The standard deviation of packet latency. Low jitter is crucial for HFT.

3.Control-Plane Convergence Time (): The time taken for the entire EVPN fabric to establish or re-establish forwarding state after an event (e.g., a link failure or a client IGMP join/leave).

3. Results and Comparative Evaluation

3.1. Scaling Limitations of Ingress Replication

The primary bottleneck for IR is the control plane overhead associated with BGP EVPN Type-6 (SMET) routes.

3.1.1. Quantitative Analysis of State Proliferation

The Type-6 route is essential for selective IR. In an environment with VTEPs and active multicast streams, the number of Type-6 routes that must be advertised by the receiving VTEPs and learned by the source VTEPs rapidly explodes. If every VTEP is interested in every group (a worst-case but common scenario in HFT):

For a moderate-sized HFT colocation with VTEPs and active streams, the BGP control plane must distribute and maintain approximately Type-6 routes. This massive route count creates several failures:

- BGP Process Strain: BGP peering sessions become burdened, leading to slow processing times and increased.
- State Overload: Every VTEP must store the forwarding list associated with these routes, consuming precious hardware resources.

3.1.2. Data Plane Overhead

While the control plane struggles, the data plane of IR also creates a bandwidth bottleneck. For a single feed (1:N fan-out) being replicated by a source VTEP to receivers, copies of the data are sent over the underlay. This over-subscription of spine bandwidth can quickly lead to buffer overflow and queuing delays, manifesting as high and unpredictable, especially during market bursts.

3.2. Performance of PIM-Integrated Solutions

PIM integration successfully solves the data-plane efficiency issue: traffic is replicated only once across the underlay, minimizing bandwidth usage and improving overall throughput. However, this approach merely shifts the state burden from the EVPN control plane to the PIM underlay and the VTEP's internal state management.

3.2.1. Complexity and State Synchronizatio

The VTEP must act as the gateway between the L2 multicast group (VNI) and the L3 PIM group (underlay). This requires:

1.PIM State Maintenance: The VTEP must maintain a large number of or entries for the PIM protocol running in the underlay, contributing to TCAM exhaustion .

2.EVPN MT Management: The Type-3 IMET route must correctly signal the use of the Multicast Tunnel (MT). Any mismatch or delay in synchronizing the EVPN state with the PIM state can cause packets to be dropped or misrouted, severely impacting.

3.Multihoming Challenges: In modern architectures, VTEPs are often multihomed to servers (e.g., using ESI or vPC). Integrating PIM with EVPN multihoming introduces complex Designated Forwarder (DF) election logic for BUM traffic, which, if slow to converge, contributes to high and data plane outages.

3.2.2. Hardware Resource Exhaustion (TCAM)

Regardless of whether the state is BGP EVPN (IR) or PIM (integrated), the ultimate performance bottleneck is the size of the hardware forwarding tables (TCAM). Every unique L2 multicast destination address or L3 multicast group requires a physical entry in the TCAM for high-speed lookup and forwarding decision. Since HFT demands dedicated, specialized network hardware (ASICs) optimized for raw packet throughput, these devices often have less TCAM depth compared to general-purpose routers.

The large number of states required for feeds in a multi-tenant environment (where each VNI compounds the state) will inevitably lead to TCAM resource exhaustion. Once TCAM is full, the device must resort to slower software processing or fail to forward the flow, leading to catastrophic failure and unpredictable latency.

4. Discussion and Proposed Enhancement

4.1. Critical Synthesis of Architectural Trade-offs

A critical synthesis of the two standard EVPN multicast forwarding mechanisms confirms a fundamental limitation in addressing the combined MGC and Jitter requirements of HFT colocation.

Feature	Ingress Replication (IR)	PIM-Integrated Underlay	HFT Requirement
Data Plane Efficiency	Very Low (Bandwidth Intensive)	High (Bandwidth Efficient)	High
Control Plane Complexity	High (State Proliferation -)	High (State Sync. and Protocol Stack)	Low
Multicast Group Capacity (MGC)	Limited (High Type-6 TCAM Burn)	Limited (High PIM TCAM Burn)	Very High (groups)
Latency Jitter ()	High (Due to bandwidth/buffer contention)	Unpredictable (Due to control-plane sync. issues)	Very Low
Operational Simplicity	High (No underlay multicast)	Low (Requires underlay PIM)	High

The results clearly show that neither standard architecture is suitable. IR fails due to bandwidth and MGC limits, while PIM-integrated solutions fail due to operational complexity and unpredictable jitter. The path forward necessitates a departure from the fully distributed control plane model.

4.2. Novel Approaches for Control-Plane Optimization

To achieve the necessary scaling, the architectural focus must shift to minimizing the forwarding state programmed into the hardware VTEPs. This requires advanced techniques for controlling the flow of BGP EVPN routing information.

4.2.1. Leveraging Network Virtualization and Route Filtering

One immediate, low-hanging optimization involves aggressively applying route filtering. While all VTEPs in the fabric must know the presence of all other VTEPs (via Type-3 IMET routes), they do not need to learn the Type-6 SMET routes for streams that neither originate from nor terminate on them. Network operators can use BGP filtering policies to only allow the Type-6 routes to propagate to VTEPs that have local receivers. While this is an improvement over full mesh distribution, it still requires the central Route Reflectors (RRs) and the spine VTEPs (which often act as RRs) to process and store the complete, exploded state

4.2.2. Application-Layer Multicast and Event-Driven Architecture

An alternative solution shifts the problem away from the network layer entirely by using application-layer multicast or event-driven architecture (EDA) principles. Instead of relying on IP Multicast for market data delivery, firms use highly optimized software components (e.g., Kafka, proprietary messaging buses) that tunnel traffic across the EVPN overlay via unicast or highly segmented multicast groups. The application logic handles the fan-out and replication. While this solves the network scaling problem, it shifts the latency and jitter burden to the application servers, requiring specialized low-latency computing environments.

4.3. Advanced Control-Plane Optimization: Leveraging Software-Defined Networking for Multicast Group Capacity Enhancement (MGC)

The preceding analysis established that standard VXLAN/BGP EVPN mechanisms—both Ingress Replication (IR) and PIM-integrated underlays—yield unacceptable trade-offs between Multicast Group Capacity (MGC) and Data-Plane Latency Jitter for demanding High-Frequency Trading (HFT) colocation environments. The primary constraints are the proliferation of BGP EVPN Type-6 (SMET) routes in the control plane and the eventual exhaustion of Ternary Content-Addressable Memory (TCAM) in the hardware forwarding plane.

To break these architectural limits, a paradigm shifts from fully distributed, stateful control towards a centralized, optimized model is required. Software-Defined Networking (SDN) offers the necessary decoupling and programmability to abstract the network state, thereby enabling advanced techniques for MGC scaling.

4.3.1. The Necessity of Decoupling: Control-Plane Offloading in HFT

In traditional distributed control planes, every device (VTEP) must maintain state for every multicast group it is interested in, or for every VTEP it needs to replicate to (in the case of IR). In an HFT colocation with VTEPs, multicast groups, and VNIs, the state complexity approaches, quickly overwhelming even high-end silicon.

SDN fundamentally addresses this by decoupling the control plane (the SDN Controller) from the data plane (the VTEPs) . This separation allows the control plane to hold the master state for the entire network, while injecting only the minimal required forwarding state into the data plane devices. This approach achieves control-plane offloading, moving the complexity burden from hardware-constrained network devices to centralized, horizontally scalable compute resources.

For multicast in particular, the controller can achieve:

1.Global View: Precise knowledge of all multicast sources, all receivers, and the complete network topology.

2.Optimized Path Calculation: Determination of the shortest, most efficient, and lowest-latency replication trees (SPT or shared trees) without relying on distributed protocols like PIM, which inherently add setup delay and state complexity.

3.State Minimization: Implementation of techniques to reduce the number of BGP EVPN routes advertised and programmed into the VTEP hardware.

4.3.2. SDN Architecture for Centralized EVPN State Management

A viable SDN architecture for this use case involves a dedicated, cluster-based controller residing outside the data path. This controller performs several key functions:

- BGP EVPN Peer: The controller peers with all Boundary VTEPs (Spines) to receive and process all BGP EVPN routes, including the critical Type-3 (Inclusive Multicast Ethernet Tag - IMET) and Type-6 (Selective Multicast Ethernet Tag - SMET) routes.
- IGMP Snooping Proxy: The controller functions as a virtual IGMP Snooping agent. Instead of relying on VTEPs to distribute IGMP join/leave messages across the network (which would generate Type-6 routes), the VTEPs simply forward all IGMP state changes to the centralized controller.
- Policy Engine: Based on the HFT client's specific subscription profiles (e.g., "Client A needs NASDAQ Equity Feeds VNI 1001"), the policy engine dictates which forwarding state is necessary.

This architecture fundamentally alters the role of the VTEP. The VTEP is reduced from a decision-making node into a simple, high-speed packet-forwarding engine, programmed entirely by the centralized intelligence of the SDN controller .

4.3.3. Mechanism 1: Selective BGP EVPN Route Advertisement (SRA)

The most direct way to enhance MGC is to minimize the programming of unnecessary state, achieved via Selective BGP EVPN Route Advertisement (SRA). This technique specifically targets the explosion of the Type-6 SMET route, which is generated for every (Source, Group, VNI) tuple subscribed to by a VTEP.

4.3.3.1. Problem with Standard Type-6 Route Advertisemen

In a standard EVPN deployment using BGP as the control plane, if a VTEP is interested in an stream, it sends a Type-6 route advertising its interest. Every other VTEP that is a source for must learn this route to establish the replication path (assuming IR). If there are groups and VTEPs, the control plane must maintain routes. This is compounded in a multi-tenant environment.

4.3.3.2. SRA Implementation via SDN

The SDN-based SRA mechanism works as follows:

1.Global State Collection: The SDN controller, acting as a BGP Route Reflector (RR) or full mesh peer, receives and stores all Type-6 routes from all VTEPs. The controller maintains a complete, persistent map of all interests.

2.Path Optimization: When a source VTEP () sends traffic for, the controller consults its global state. It identifies only the specific destination VTEPs () that have active receivers for that exact stream.

3.Selective Injection: The controller dynamically creates and advertises a filtered subset of Type-6 routes only to the relevant source VTEPs () that need to perform replication, and only with the next-hop information of the currently active subscribers.

This targeted approach drastically reduces the state programmed on the VTEPs. Only the specific VTEPs involved in traffic replication for active streams hold the necessary forwarding state, leading to two major benefits:

- TCAM Conservation: VTEPs, particularly those acting as sources or transient replication points, conserve valuable TCAM resources, allowing the platform to support a significantly higher MGC than otherwise possible.
- Reduced Control-Plane Traffic: The overall volume of BGP updates is reduced, enhancing control-plane stability and convergence speed during link-up/link-down events.

4.3.4. Mechanism 2: Hierarchical Multicast Abstraction (HMCA) and Route Aggregatio

While SRA minimizes state on a per-VTEP basis, Hierarchical Multicast Abstraction (HMCA) seeks to minimize the number of unique BGP routes advertised for multiple related streams by aggregating them logically. This technique is especially powerful in HFT, where market data is often structured into tiered feeds (e.g., Level 1, Level 2, Full Depth) across hundreds of different instruments, all belonging to the same VNI or a small set of VNIs.

4.3.4.1. The Need for Group Aggregation

In financial feeds, streams are often logically grouped. For example, all equities feeds for the New York Stock Exchange might reside in VNI 200, with thousands of different multicast groups. Standard EVPN requires a separate Type-6 route for every single combination, leading to the issue.

HMCA introduces a hierarchical approach by defining a Meta-Multicast Group (MMG) that encapsulates a collection of individual streams.

4.3.4.2. Implementing HMCA via Type-3 IMET Optimization

Instead of relying solely on Type-6 (SMET) per-stream routes, HMCA leverages an optimization of the Type-3 (IMET) route, typically used for unknown multicast traffic or Broadcast/Unknown Unicast (BUM) traffic.

1.MMG Definition: The SDN controller pre-defines an MMG—a single, generic L2 Multicast Group (e.g.,) associated with the VNI.

2.IMET Advertisement: When a source VTEP sends data for any stream within the defined MMG/VNI, the controller uses the Type-3 IMET route to establish the initial distribution tree to all interested VTEPs. This is done once per VNI, rather than once per .

3.Application-Layer Filtering: This shifts the filtering responsibility to the edge. The VTEP forwards the encapsulated MMG traffic to the receiver host. The host application then performs the final packet filtering based on the Layer 4 (UDP) port or application-layer header, discarding streams it did not subscribe to.

While this introduces an element of over-replication (sending the MMG traffic to VTEPs who may not need all streams within the MMG), the trade-off is often acceptable in dedicated HFT colocation environments because:

- Control-Plane Simplification: The controller only needs to manage state for the far smaller number of MMGs (e.g., 50 MMGs instead of 50,000 unique streams).
- Predictable Jitter: While the data plane handles slightly more traffic, the control-plane stability is vastly improved, minimizing unpredictable Latency Jitter caused by control-plane convergence events, which is arguably a more critical metric in HFT than raw throughput.

The implementation of HMCA requires the HFT applications to be optimized for receiving and filtering aggregated feeds, an architectural shift that is increasingly common with event-driven architectures.

4.3.5. Comparative Modeling of MGC Improvement and Latency Impact

To quantitatively demonstrate the effectiveness of SRA and HMCA, we can model the relative change in MGC (measured by the percentage reduction in programmed TCAM entries) and the resulting impact on convergence and jitter.

4.3.5.1. MGC Improvement Modeling

We use the following parameters, based on current industry benchmarks for HFT colocation:

- unique multicast groups (feeds).
- VTEPs (Leaf Switches).
- total usable multicast entries per VTEP.
- source VTEPs sending a given stream.
- VTEPs receiving a given stream.

Mechanism	State Complexity (Per Stream)	Total TCAM Entries Required	MGC (Theoretical)	MGC % Improvement over Standard EVPN
Standard EVPN IR		(Controller)	groups	N/A
SRA (SDN-based)		(Controller only)	(VTEP uses flow)	**~400%**
HMCA (Type-3 IMET)			Limited by for	**Depends on Aggregation Ratio**

Note on SRA TCAM: While the controller manages the full state, the VTEP only needs one entry per VNI/multicast interface to point to the next-hop VTEPs. If the multicast replication list is entirely handled by the SDN controller, the TCAM burden is significantly reduced, potentially freeing up capacity by a factor of where is the number of Meta-Multicast Groups.

4.3.5.2. Jitter and Convergence Impact

The trade-off for the MGC scaling must be analyzed against the crucial HFT metric of Latency Jitter.

- SRA Impact on Jitter: The major jitter contributor is the delay between an IGMP join/leave event (client request) and the corresponding Selective BGP Route Injection (SRI) by the SDN controller.
- Convergence Time ():.
- To maintain ultra-low latency, the combined processing and injection time () must be measured in sub-millisecond intervals (e.g.,). This mandates the use of highly optimized, often in-memory database solutions and low-level protocol interactions (e.g., gRPC or P4-enabled programming) for the controller-VTEP link.
- HMCA Impact on Jitter: HMCA essentially stabilizes the control plane by reducing the number of dynamic routes. Since the major replication paths (the Type-3 IMET routes) are semi-static and pre-programmed, this mechanism eliminates the jitter associated with frequent Type-6 route withdrawals and advertisements. The primary impact is a slight, constant increase in data-plane latency due to the host application's need to filter the extra, unnecessary MMG packets—a predictable overhead that is generally preferred over unpredictable control-plane jitter.

4.3.6. Security and Compliance Implications of SDN-Controlled State

Moving to a centralized, SDN-controlled architecture introduces new considerations, particularly in the highly regulated financial sector .

- Single Point of Failure (SPOF): The SDN controller cluster becomes a critical SPOF. Resilience is non-negotiable, requiring a geographically dispersed, highly available (HA) cluster with robust data consistency protocols to ensure the global network state remains synchronized.
- Security Posture: Since the controller is managing the entire flow state, it is a high-value target for denial-of-service (DoS) or unauthorized state injection. The control plane must be isolated, authenticated, and secured using methods like Transport Layer Security (TLS) for all controller-to-VTEP communication.
- Regulatory Compliance: Any architectural change must satisfy audit requirements for data integrity and network performance guarantees. The centralized log of all IGMP joins/leaves and corresponding SRA injections provides an excellent audit trail, potentially superior to tracing distributed protocol logs across dozens of VTEPs. This is a crucial advantage for maintaining compliance records.

In summary, leveraging SDN to implement Selective Route Advertisement (SRA) and Hierarchical Multicast Abstraction (HMCA) represents the most promising path forward for achieving the requisite MGC in next-generation VXLAN/BGP EVPN HFT colocations. The transition requires a move toward high-performance, resilient, and specialized controller logic that can operate within the stringent sub-millisecond convergence constraints dictated by modern algorithmic trading.

4.4. Limitations and Future Research Directions

4.4.1. Study Limitations

The primary limitation of this study is its foundation in analytical modeling and simulation-based evaluation. Real-world deployment data from major HFT exchange fabrics remains proprietary and inaccessible for public validation. Consequently, the quantitative MGC improvement figures are theoretical maximums that assume perfect control-plane synchronization and ideal VTEP hardware behavior. Further, the model does not account for the non-linear increase in data-plane latency associated with the small but necessary over-replication inherent in HMCA.

4.4.2. Future Direction 1: Experimental Validation of SDN Mechanisms

The most immediate future research should focus on experimental validation of the proposed SRA and HMCA mechanisms. This requires building a high-fidelity testbed using programmable network hardware (e.g., based on P4 language or merchant silicon) and deploying an open-source or proprietary SDN controller capable of sub-millisecond route injection. The goal would be to empirically measure and under extreme load to confirm the practical MGC limits and the latency trade-offs of the centralized model

4.4.3. Future Direction 2: Investigating ICN and NDN Overlays

Beyond optimizing the existing EVPN/IP framework, future research should investigate fundamental shifts in the underlying networking paradigm. Emerging architectures, such as Information-Centric Networking (ICN) or Named Data Networking (NDN), naturally support content-based routing and multicast by design. These overlays, which prioritize content retrieval over address-based routing, could inherently solve the state explosion problem by eliminating the need for explicit state, replacing it with content-name state. The applicability and performance of NDN overlays to the specific requirements of low-latency financial feeds warrant detailed study.

5. Conclusion

The standard implementation of VXLAN/BGP EVPN, while a robust platform for general multi-tenant data center virtualization, proves insufficient for the highly demanding multicast scaling and low-jitter requirements of High-Frequency Trading colocation environments. The distributed control plane mechanisms of both Ingress Replication and PIM-integrated solutions lead to an intractable Multicast Group Capacity (MGC) problem due to the explosion of required BGP EVPN Type-6 routes and subsequent hardware TCAM resource exhaustion.

To overcome these limitations, a shift toward a Software-Defined Networking (SDN) driven architecture is essential. By implementing advanced control-plane optimizations, specifically Selective Route Advertisement (SRA) and Hierarchical Multicast Abstraction (HMCA), the state burden can be offloaded from VTEP hardware to a centralized, scalable controller. This decoupling allows for a significant increase in MGC while stabilizing the control plane, thereby reducing unpredictable latency jitter—the most critical performance metric in HFT. Future research must empirically validate these theoretical performance gains and explore disruptive networking technologies like NDN to ensure the continuous evolution of financial network infrastructure.

The modernization of trading colocation infrastructures demands not only optimized multicast performance but also rigorous security validation frameworks to safeguard network reliability. According to Kumar Tiwari (2023), the implementation of automated security testing within digital transformation ecosystems ensures real-time detection of vulnerabilities and enhances system resilience against evolving cyber threats. Such automation principles can be extended to VXLAN/BGP EVPN environments, where low-latency network functions must operate securely under continuous integration and deployment cycles.

References

1. Chandra Jha, A. (2025). VXLAN/BGP EVPN for trading: Multicast scaling challenges for trading colocations. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3478
2. 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
3. Singh Chadha, K. (2025). Edge AI for real-time ICU alarm fatigue reduction: Federated anomaly detection on wearable streams.

Utilitas Mathematica, 122(2), 291–308. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2708

4. Ghosh, S. (2023). Building Low Latency Applications with C++: Develop a complete low latency trading ecosystem from scratch using modern C++. Packt Publishing Ltd.
5. 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.2015.244010
6. 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-drivenarchitecture-microservices-patterns-pitfalls-andbest-practices
7. Nagaraj, V. (2025). Ensuring low-power design verification in semiconductor architectures. Journal of Information Systems Engineering and Management, 10(45s), 703–722. https://doi.org/10.52783/jisem.v10i45s.8903
8. Alshammari, A. R. (2020). Resilient Wireless Network Virtualization with Edge Computing and Cyber Deception (Doctoral dissertation, Howard University).
9. 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
10. Balbaa, M. E. (2022). International Transport Corridors. Tashkent State University of Economics: Tashkent, Uzbekistan.
11. 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
12. 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
13. 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
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. Hariharan, R. (2025). Zero trust security in multi-tenant cloud environments. Journal of Information Systems Engineering and Management, 10(45s). https://doi.org/10.52783/jisem.v10i45s.8899
16. Reddy Gundla, S. (2025). PostgreSQL tuning for cloud-native Java: Connection pooling vs. reactive drivers. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3479
17. Samantapudi, R. K. R. (2025). Enhancing search and recommendation personalization through user modeling and representation. International Journal of Computational and Experimental Science and Engineering, 11(3), 6246–6265. https://doi.org/10.22399/ijcesen.3784
18. 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
19. 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-schedulingimproving-patient
20. 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

21. Enugala, V. K. (2025). "BIM-to-field" inspection workflows for zero paper sites. Utilitas Mathematica, 122(2), 372–404. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2711
22. George, J. (2022). Optimizing hybrid and multicloud 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
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. Blanchard, D. (2021). Supply chain management best practices. John Wiley & Sons.
25. 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
26. 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/THECONVERGENCE-OF-PREDICTIVEANALYTICS-IN-DRIVING-BUSINESSINTELLIGENCE-AND-ENHANCING-DEVOPSEFFICIENCY.pdf
27. 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
28. 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
29. 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
30. 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
31. Chadha, K. S. (2025). Zero-trust data architecture for multi-hospital research: HIPAA-compliant unification of EHRs, wearable streams, and clinical trial analytics. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3477
32. 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
33. Tafreshi, V. H. F. (2015). Secure and robust packet forwarding for next generation IP networks. University of Surrey (United Kingdom).
34. 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-schedulingimproving-patient
35. 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

36. Morel, L. P. (2017). Using ontologies to detect anomalies in the sky. Ecole Polytechnique, Montreal (Canada). https://www.proquest.com/openview/1310c97e55ee11adc005c478ad646164/1?pqorigsite=gscholar&cbl=18750
37. Gannavarapu, P. (2025). Performance optimization of hybrid Azure AD join across multi-forest deployments. Journal of Information Systems Engineering and Management, 10(45s), e575–e593. https://doi.org/10.55278/jisem.2025.10.45s.575
38. Chen, L. (2017). Performance Evaluation for Secure Internet Group Management Protocol and Group Security Association Management Protocol (Doctoral dissertation, Concordia University). https://libraryarchives.canada.ca/eng/services/serviceslibraries/theses/Pages/item.aspx?idNumber=1135022369
39. 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
40. 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
41. 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.2024.3426094
42. 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
43. 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
44. Sayyed, Z. (2025). Application-level scalable leader selection algorithm for distributed systems. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3856
45. 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
46. Gaurav Malik. (2025). Integrating Threat Intelligence with DevSecOps: Automating Risk Mitigation before Code Hits Production. Utilitas Mathematica, 122(2), 309–340. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2709
47. Kumar Tiwari, S. (2023). Security testing automation for digital transformation in the age of cyber threats. International Journal of Applied Engineering & Technology, 5(S5), 135–146. Roman Science Publications.