Universal Library of Engineering Technology Review Article

ISSN: 3064-996X | Volume 3, Issue 1

Open Access | PP: 22-28

DOI: https://doi.org/10.70315/uloap.ulete.2026.0301004

The Effectiveness of Combining VXLAN, BGP, and EVPN in Large-Scale Networks for Multi-Tenant Data Centers

Aghasi Gevorgyan

Head of Network Infrastructure, Armenian Card CJSC.

Abstract

The article examines the effectiveness of combining VXLAN, BGP, and EVPN technologies in designing large-scale, multi-tenant data centers with predominantly east–west traffic. The objective of the study is to provide a conceptual and analytical substantiation of a reference data-center fabric architecture that separates the L3 underlay from the overlay plane, is capable of scaling segmentation to millions of logical domains, ensures strict tenant isolation, and supports automatable workload mobility. The relevance of the research is driven by the growth of microservice and cloud workloads, which turn classical L2 extension and VLAN-based segmentation into a source of systemic unpredictability due to flooding and the limited identifier space. The scientific novelty of the work lies in interpreting the VXLAN–EVPN–BGP combination as an integral, controllably scalable system in which the 24-bit VNI, the distributed gateway, and the protocol-formalized dissemination of reachability information form a stable invariant of multi-tenant segmentation, as well as in introducing a diagnostic matrix that operationalizes underlay/overlay separation when identifying the causes of degradations. It is shown that such an architecture reduces the share of parasitic broadcast and unknown unicast traffic, stabilizes latency for inter-server exchange, localizes failure domains in inter-site connectivity, and increases infrastructure readiness for template-driven automation of configurations and policies. The article will be useful to data-center network architects, operations engineers, and developers of cloud-based educational platforms in a multi-tenant environment.

Keywords: *VXLAN, BGP, EVPN, Multi-Tenant Data Center, Virtual Segmentation.*

INTRODUCTION

Within modern large multi-tenant datacenters, which have become the basic substrate of cloud computing, and virtualization and microservice architectures where the compute fabric is dynamically tailored around workload and isolation needs, the traffic profile is dominated by internal east-west traffic, since distributed applications are composed of many small independent components that communicate with each other, rather than communicating with a few components outside the datacenter perimeter. This dominance is explicitly documented in research on large-cloud networking, which emphasizes that the primary volume of flows proceeds in the server-to-server direction within the data center (Zhang et al., 2022).

The growth of internal traffic intensifies architectural requirements: segmentation must be scaled while simultaneously enforcing strict tenant isolation and preserving predictable latency under dynamic workload migration. The classical model, extends Layer 2, keeps Layer 3 at the boundary, fits these conditions increasingly poorly, because multi-tenancy drives an avalanche-like increase in network domains and policies, while microservice architectures raise the number of short, frequent, interdependent sessions. Consequently, engineering practice and standardization have converged on separating the connectivity transport (a simple IP fabric) from the logic of tenant virtual networks, moving the latter into tunnel-based virtualization and a managed control exchange (Li et al., 2021).

Traditional L2/L3 approaches encounter scaling and resilience constraints. VLAN-based segmentation has a strict identifier ceiling, while large Layer 2 domains are compelled to rely on STP and learning mechanisms driven by observed frame traversal, which increases the risk of broadcast bursts and degradation under uncertain delivery (flooding) of unknown unicast traffic. These issues are explicitly described in the VXLAN standard as motivations for introducing the 24-bit segment identifier (VNI), which enables scaling virtual networks to millions and constructing them over IP transport without inflating Layer 2 domains (IETF, 2020). Additionally, contemporary surveys of data-center networks

Citation: Aghasi Gevorgyan, "The Effectiveness of Combining VXLAN, BGP, and EVPN in Large-Scale Networks for Multi-Tenant Data Centers", Universal Library of Engineering Technology, 2026; 3(1): 22-28. DOI: https://doi.org/10.70315/uloap.ulete.2026.0301004.

note that the complexity of traffic management and behavioral predictability increases as a systemic problem with growing scale and application dynamism, making unmanaged broadcast and flat Layer 2 domains particularly costly to operate (Liu et al., 2025).

Despite extensive research on individual data center networking technologies, including overlay encapsulation and traffic management mechanisms, insufficient attention has been paid to the systemic effectiveness of their combined operation as a coherent architectural solution. In particular, the interaction between underlay/overlay separation, control-plane formalization, and operational diagnostics in large multi-tenant environments remains underexplored. This gap motivates the present study, which analyzes the VXLAN–EVPN–BGP combination not as a set of isolated tools, but as an integrated fabric architecture aligned with scalability, isolation, and operational predictability requirements.

MATERIALS AND METHODOLOGY

The research materials were assembled as a purposive corpus of 8 sources reflecting three complementary strata: (1) empirical and engineering observations on the dominance of east–west traffic in large clouds and the associated requirements for inter-server communication (Zhang et al., 2022); (2) survey works describing the evolution of traffic management in data-center networks from the link layer to the application layer and articulating the underlay/overlay separation motif as a response to increasing dynamism and multi-tenancy (Li et al., 2021), as well as systematizing challenges of load balancing and resilience when scaling fabrics (Liu et al., 2025); (3) normative technical documents defining formal mechanisms for overlay segmentation and the control plane: RFC 7348 for VXLAN (IETF, 2020) and RFC 8365 for the EVPN overlay (RFC, 2018). The corpus additionally includes applied studies on integrating VXLAN into infrastructure practice and achieving scalable segmentation via the 24-bit identifier (Efendi et al., 2023), as well as a work representing the class of dynamic cloud resource allocation problems as a context that strengthens requirements for workload mobility and automation (Ali et al., 2025).

Methodologically, the work is conducted as a conceptual-analytical study with a systematized comparison of architectural patterns. In the first step, a thematic analysis was conducted to identify factors that render traditional L2 extension operationally fragile under growth in the number of domains and the microservice granularity of traffic, based on observations of flow structure and inter-server link load (Zhang et al., 2022; Li et al., 2021). In the second step, a protocol-semantic analysis was carried out for VXLAN and EVPN as a coupling of data plane and control plane: VXLAN was treated as an encapsulation mechanism and scalable segment identification via VNI, while EVPN was treated as a reachability dissemination model and MAC/IP binding distribution that reduces dependence on flooding; the formal basis for these interpretations was extracted from the relevant RFC descriptions (IETF, 2020; RFC, 2018). In the third step, the results were mapped to the requirements of a large multi-tenant data center (scalable segmentation, isolation, predictable latency, resilience, and suitability for automation), where emphases on multipath delivery and overload behavior were refined through a load-balancing survey (Liu et al., 2025), and the practical applicability of VXLAN-oriented approaches was grounded via an applied case of integration and configuration automation (Efendi et al., 2023).

To operationalize the conclusions, a synthetic design–artifact approach was used: a reference model L3 underlay as transport + overlay as segment semantics and policy was formed, and functions were then explicitly decomposed across planes to render verifiable the causal relationships symptom, layer, invariant (Li et al., 2021; IETF, 2020). As an instrument for verifying internal consistency (not empirical, but logical-structural), a diagnostic matrix of observable deviations was constructed, linking loss/jitter/endpoint disappearance and broadcast spikes to underlying pathologies (ECMP, micro-congestion, line quality) and to errors or staleness of overlay state (reachability advertisements, excessive flooding), which follows directly from the role of the EVPN control plane and the nature of overlay encapsulation (RFC, 2018; Liu et al., 2025). A methodological limitation is that the work deliberately does not introduce laboratory measurements; instead, it demonstrates the effectiveness of the combination by engineering feasibility and controllable scalability through the alignment of requirements, protocol mechanisms, and operational procedures (Efendi et al., 2023; Ali et al., 2025).

RESULTS AND DISCUSSION

Following the limitations of traditional solutions outlined in the introduction, the key requirements for the network of a large multi-tenant data center can be summarized as managed growth of complexity without loss of predictability. The first requirement is segmentation scalability: the number of logical networks for tenants and applications rapidly exceeds the scale of thousands and tends toward the order of millions if segmentation is constructed not only by organizations, but also by environments, access zones, and data classes. This is precisely why modern architectures rely on a 24-bit segment identifier, enabling up to 16 million logical domains over an underlying Layer 3 network without inflating physical switching itself (Efendi et al., 2023).

The second requirement is tenant isolation and security. This acquires a direct applied meaning: different educational organizations, courses, roles, and content storage domains must not intermingle in addressing, routing, or access policy. Systematic reviews of multi-tenancy in cloud-oriented architectures emphasize that isolation is not a single

configuration, but a set of mechanisms across multiple layers, and a weak link can turn co-location into a source of cascading risks to confidentiality and resilience (Ali, Talpur, et al., 2025). Therefore, the network must support separate virtual routed domains, managed inter-domain interaction, and reproducible policies that are interpreted identically across all fabric nodes.

The third requirement is fault tolerance and predictable latency, because dense internal east–west traffic is sensitive to micro-congestion and path asymmetry; small fluctuations translate into noticeable degradation at the application and educational platform levels. Contemporary surveys on load balancing in data-center networks indicate that increasing scale and workload complexity make effective balancing and resilience critical QoS conditions; accordingly, the architecture should natively employ multipath routing, fast connectivity restoration, and controlled dissemination of reachability information (Liu et al., 2025).

The fourth requirement is workload mobility and automation support, because compute components (virtual machines, containers, services) move and scale faster than manual network re-carving can accommodate. Research on data-center fabric virtualization highlights the significance of software-defined allocation and modification of network resource slices to accommodate application requirements, including dynamic expansion and contraction in response to load changes (Ali et al., 2025). In a multi-tenant educational environment, this directly corresponds to the need to rapidly enable new learning streams and isolated domains without downtime and prolonged approvals. Essential Network Requirements are shown in Figure 1.

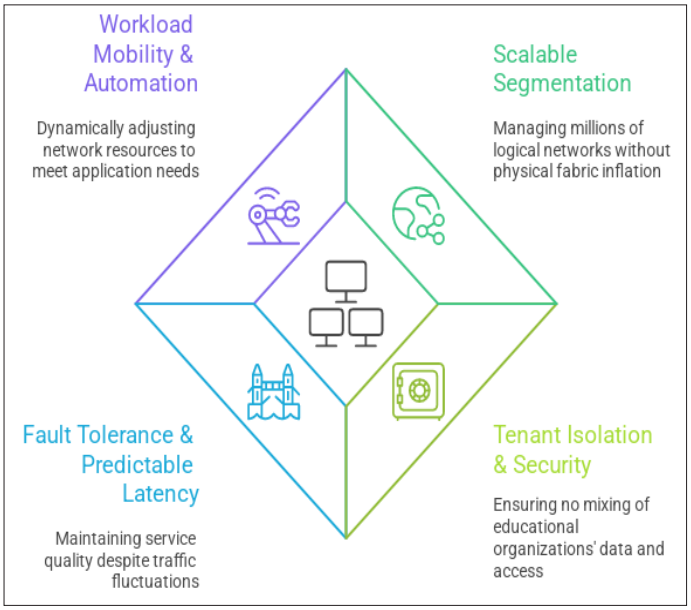

Fig. 1. Essential Network Requirements

Against this background, the role of virtual LAN extension technology is to implement an overlay network in which Layer 2 frames are delivered over a Layer 3 underlay by encapsulation in user datagram packets, while segment membership is determined by a network identifier. Tunnel endpoints at the boundary of compute racks perform encapsulation and decapsulation, allowing the physical fabric to be built as a simple routed network, while logical domains are realized as manageable segments at a scale unattainable for classical virtual LANs.

The control plane in such an architecture must replace learning via unknown-traffic flooding with targeted dissemination of information about endpoints and their addresses. This is enabled by Layer 2 VPN technology, in which reachability information for link-layer and network-layer addresses is distributed as routes, while also supporting a distributed gateway model with a single address that shortens paths and stabilizes latency for internal exchanges. The standard describing the application of this technology to overlay networks explicitly addresses its joint operation with encapsulation and defines procedures that reduce the share of broadcast and unknown unicast traffic and simplify multi-tenant segmentation (RFC, 2018).

Finally, the Border Gateway Protocol serves as the transport for these reachability routes and simultaneously becomes a mechanism for scaling and policy: it carries information among fabric nodes, enables path preference and propagation constraints, and provides predictable convergence in the event of failures. As a result, the combination of encapsulation for the data plane, Layer 2 VPN for the control plane, and BGP for route distribution forms a coherent toolset that directly satisfies the requirements of scalability, isolation, resilience, and automability formulated for multi-tenant data centers.

The reference architecture for a multi-tenant data center is derived from the previously discussed requirements of scalability, isolation, and predictability, and is therefore constructed as a fabric with a simple transport substrate and managed virtual network logic above it. As the physical framework, a fabric with spine and leaf tiers is used: leaf nodes connect compute servers, and spine nodes provide non-blocking multipath connectivity between racks. The meaning of this geometry lies not in diagram aesthetics, but rather in the transformation of the network into a reproducible module, where adding new racks minimally alters the routing logic and does not create hidden dependencies between domains.

The underlay is built as a Layer 3 transport, prioritizing resilience and simplicity over excessive functionality. Typically, interior routing protocols or border routing between neighbors are selected, as they provide strong support for multipath forwarding and fast failure recovery. Configuration requirements are minimized, ensuring that the underlay remains predictable even as scale increases and service-level changes occur frequently.

The overlay solves the opposite task: not merely to deliver a packet, but to preserve the semantics of a virtual segment

for the tenant and application. For this, Layer 2 frames are encapsulated into transport packets of the underlay, while segment membership is encoded by a separate identifier that is not constrained by the tight limits of classical local segmentation. Tunnel endpoints are placed on leaf nodes and serve as the boundary where traffic is translated from the virtual network domain into the transport fabric and back.

The interaction of tunnel endpoints is important not only in the data plane, but also in the control plane: the network must know in advance where an endpoint resides so that it does not flood traffic in the hope that it will eventually find a recipient. This is where the coupling of BGP and a Layer 2 VPN technology comes into play, converting reachability into routable advertisements and thereby rendering network behavior more deterministic.

Control-plane exchange typically includes advertisements of link-layer addresses and associated network-layer addresses, information on segment membership, and parameters required for the correct operation of the distributed gateway with a single address. As a result, the first hop for compute nodes becomes local, while subsequent forwarding selects an optimal path through the fabric without forcing flows to hairpin through a centralized routing point, which is especially critical under intensive inter-server exchange.

Special attention is required for the delivery of broadcast, unknown unicast, and multicast frames, as they become a source of noise and unpredictability in classical Layer 2 domains. In an overlay network, there are two typical methods: organize multicast distribution through the transport fabric, or perform ingress replication, where the leaf node generates multiple copies and sends them unicast to the required tunnel endpoints. The choice between these approaches typically depends on hardware capabilities and on which objective is more important in a particular data center: saving state in the network or saving bandwidth at the rack nodes.

For the combination to function as an engineering system rather than a set of tricks, the separation-of-responsibility principle must be stated explicitly: the underlay is responsible for connectivity and recovery, while the overlay is responsible for segment semantics, isolation, and policy. Table 1 illustrates the distribution of functions between the underlay and overlay parts of the network.

Table 1. Distribution of functions between the underlay and overlay parts of the network

Area	Underlay (Layer 3 network)	Overlay network & control plane
Rack-to-rack connectivity	Multipath forwarding and fast recovery	Uses the connectivity as a transport
Scaling	Adding nodes without expanding Layer 2 domains	Growth in the number of segments via identifiers
Isolation	Basic filtering and path separation when needed	Virtual tenant domains and segments
Access policies	Minimal routing required	Inter-segment communication rules and constraints
Diagnostics	Verifying paths, loss, and latency	Verifying reachability advertisements and endpoint mappings

From a system perspective, the results indicate that the effectiveness of the VXLAN–EVPN–BGP combination is not reducible to any single protocol. Rather, it emerges from the explicit separation of responsibilities between transport and semantics, where the underlay guarantees connectivity and convergence, while the overlay formalizes segmentation, isolation, and policy. This separation constrains failure propagation and renders network behavior more predictable under growth and change.

Within a single site, the effectiveness of the combination primarily manifests in reducing the spill of unknown traffic: instead of learning endpoint locations through broad dissemination, the network obtains this knowledge through control-plane advertisements. As a result, parasitic load decreases, latency becomes easier to keep within a narrow corridor, and growth in the number of endpoints does not produce a proportional growth in noise that would otherwise be inevitable in large Layer 2 domains.

A distributed gateway with a single address shortens and stabilizes inter-server paths because the first hop is executed locally on the leaf node rather than somewhere in the middle. For learning platforms and assessment services, this translates into more stable response times during peak sessions, when many microservices exchange data simultaneously, and load increases sharply during mass module completion and certification events.

Growth in the number of segments becomes manageable because segmentation ceases to be a function of physical switching and becomes a logical construct that can be created and removed on demand. Practically, this means that separate domains can be allocated for organizations, learner groups, lab environments, and analytics services without changing the fundamental transport fabric scheme or expanding Layer 2 domains to unsafe sizes.

Failure resilience in such an architecture is based on two complementary mechanisms: the underlay rapidly recalculates routes, while the overlay control plane updates reachability information for endpoints. With correct configuration, this yields predictable behavior under leaf-node or link failure: traffic either quickly transitions to an

alternate path, or, if the endpoint is truly unavailable, ceases attempts to deliver frames, thereby avoiding self-sustaining congestion.

In multi-site scenarios, the principal fork lies between interconnecting fabrics at Layer 3 and stretching Layer 2 domains across an inter-site link. Layer 3 connectivity is typically simpler to operate because failure domains remain local and inter-site latency does not force Layer 2 protocols to operate in an unnatural regime; tenant semantics can still be preserved through virtual routed domains and aligned exchange policies.

Stretching Layer 2 domains is justified only where mandated by the application, for example, under a rigid dependency on a single addressing plan or under specific migration scenarios. However, the cost of such a decision is almost always higher than anticipated. Risks include expanding the broadcast domain over an inter-site channel, which complicates fault localization and increases sensitivity to loss and jitter, both of which are difficult to eliminate completely in inter-site environments, even with robust transport infrastructure.

Tenant isolation across sites requires discipline at the level of virtual domains and reachability propagation rules. What appears as a convenient shared network within a single site may become a channel for unintended domain mixing at the inter-site scale. Therefore, a reasonable recommendation reduces to keeping tenant virtual domains closed by default, enabling inter-site exchange only where required by the learning or analytics scenario, and defining in advance the control points through which permitted traffic passes.

Security policy in a multi-tenant environment naturally relies on the coupling of virtual routed domains and segment identifiers, because it turns isolation from a declaration into a technical fact. On this basis, additional segmentation can be built, up to microsegmentation, where rules are defined not only between networks but also between service roles and workload groups, which is particularly useful for systems with diverse external integrations and different access levels to materials.

When needed, inter-tenant access control can be implemented as managed traffic traversal through a chain of network services, where each step performs a constrained function, such as policy verification, filtering, or inspection. It is essential that such sequencing become a design norm rather than a set of exceptions; otherwise, the network becomes an archive of special cases, and even a robust architecture loses its advantages due to the inability to reproducibly explain why a packet was forwarded or dropped.

Operations and automation bind all prior decisions into a unified lifecycle. Configuration templates enable the onboarding of new tenants without manual fine-tuning, and change procedures become structurally comparable, even when tenants differ in scale and service portfolios. For observability, it is useful to separate monitoring of underlay and overlay components because the same application-level symptom may have different causes, and this separation accelerates diagnosis. Table 2 provides a concise inspection matrix for typical deviations, facilitating systematic troubleshooting.

Table 2. Matrix for inspection of typical deviations

Observed effect	What to check in the underlay	What to check in the overlay
Packet loss and retransmissions	Link congestion, physical line quality, and how evenly multipath forwarding is distributing traffic	Correctness of reachability information and absence of unnecessary flooding
Increased latency and jitter	Uneven utilization of paths and queues	First-hop locality via the distributed gateway
A node disappears from a segment	Rack node availability and basic connectivity	Freshness/validity of advertised link-layer and network-layer addresses
Spike in broadcast traffic	Signs of port storms and backbone/uplink saturation	Broadcast handling mode and correctness of replication
Unexpected cross-tenant access	Inter-domain routing and filters	Inter-segment communication rules and policy enforcement points

The table offers a diagnostically structured decomposition of fault symptoms across the underlay–overlay boundary, effectively operationalizing the principle of separation of concerns in multi-tenant data center fabrics. It indicates that loss, retransmissions, and latency inflation are often rooted in underlay pathologies, such as micro-congestion, impaired physical links, or uneven ECMP utilization, yet may be amplified or even misattributed when the overlay control plane disseminates stale reachability state or triggers superfluous flooding. Likewise, endpoint disappearance is framed as a coupled failure mode in which basic rack reachability is necessary but not sufficient unless MAC/IP advertisements remain timely and consistent. The broadcast-spike row highlights how L2 replication semantics can transform localized anomalies into fabric-wide stress, whereas the cross-tenant access row underscores that confidentiality breaches typically emerge from policy-plane and segmentation misconfiguration, even when the routed substrate is nominally correct. Overall, the matrix functions as an epistemic map for troubleshooting, partitioning

observables into substrate-level transport constraints and overlay-level state/policy invariants, thereby supporting more reproducible and falsifiable diagnostic workflows.

CONCLUSION

Within the considered problem space, it is demonstrated that in large multi-tenant data centers, the underlying dominance of east–west traffic and the dynamism of microservice workloads foreground not so much raw throughput as the managed complexity of architecture, while preserving latency determinism and isolation. The limitations of traditional L2 extension and VLAN-based segmentation manifest in a strict identifier ceiling and an operationally hazardous dependence on broadcast mechanisms and flooding, making the growth of Layer 2 domains a source of systemic unpredictability. Therefore, the logic of separating responsibilities, where the underlay functions as a simple L3 transport and tenant virtual networks are implemented as a managed overlay, is presented in the text not as a stylistic preference but as a response to the structural properties of modern data centers.

The formulated requirements, including scalability to the order of millions of logical domains, strict multi-level tenant isolation and policy reproducibility, fault tolerance and latency stability, as well as workload mobility and automation suitability, are unified into a single engineering contour through the combination of encapsulation, controlled dissemination of reachability information, and predictable convergence. The overlay network with a 24-bit segment identifier eliminates the dependence of logical segmentation on physical switching; the control plane replaces learning after the fact with targeted knowledge about endpoints, thereby reducing the share of broadcast and unknown unicast traffic. The distributed gateway with a single address sustains first-hop locality, stabilizing inter-server exchange behavior. Consequently, the combination of VXLAN as a forwarding mechanism, EVPN as a reachability advertisement model, and BGP as the transport for these advertisements is interpreted as an integral system directly aligned with multi-tenancy requirements.

The reference fabric geometry, as described in the discussion section, further emphasizes that scale is achieved not by inflating fragile Layer 2 domains, but by rack-level modularity and equal-cost multipath connectivity between racks, where growth in the number of nodes minimally affects the routing logic. Under inter-site expansion, the analysis reduces the key decision fork to a choice between Layer 3 connectivity and Layer 2 stretching, with the operationally simpler preference belonging to the variant that keeps failure domains local and does not carry broad dissemination across the inter-site environment; inter-site isolation in this case requires discipline in keeping virtual domains closed by default and explicitly defining control points for permitted exchange. Thus, multi-tenancy is interpreted as a continuous task of aligning segmentation, routing, and policy rather than as a single configuration action.

Finally, the operational viability of the architecture is substantiated through the principle of observable underlay/overlay separation and through the diagnostic matrix, which translates the search for degradation causes from craft into a reproducible procedure for checking transport constraints and state/policy invariants. Table 2 demonstrates that similar application-level symptoms may have causes at different layers, from micro-congestion and uneven ECMP to stale reachability advertisements and excessive flooding, implying that correct support requires symmetric monitoring of both planes.

Overall, the study's conclusion can be formulated as confirmation that the effectiveness of combining VXLAN, BGP, and EVPN primarily manifests in controllable scalability, including the reduction of parasitic traffic, formalization of isolation and policies, predictable recovery after failures, and technological readiness for automating the lifecycle of multi-tenant segments. The study contributes to modern data center networking research by framing the VXLAN–EVPN–BGP combination as an architectural invariant that enables controllable scalability, formalized isolation, and reproducible operations in multi-tenant environments, rather than as a collection of protocol-level optimizations.

REFERENCES

1. Ali, K. A., Fadare, O. A., & Al-Turjman, F. (2025). Dynamic Resource Allocation (DRA) in Cloud Computing. *Sustainable Civil Infrastructures*, 1033–1049. https://doi.org/10.1007/978-3-031-72509-8_85
2. Ali, S., Talpur, D. B., Abro, A., Alshudukhi, K. S. S., Alwakid, G. N., Humayun, M., Bashir, F., Wadho, S. A., & Shah, A. (2025). Security and privacy in multi-cloud and hybrid cloud environments: Challenges, strategies, and future directions. *Computers & Security*, *157*, 104599. https://doi.org/10.1016/j.cose.2025.104599
3. Efendi, N. A., Husna, N. D., & Nugraha, N. I. G. D. (2023). Advancing Network Infrastructure: Integrating VXLAN Technology with Automated Circuit Operations and NOS Configurations. *International Journal of Electrical, Computer, Biomedical Engineering*, *1*(2), 32–53. https://doi.org/10.62146/ijecbe.v1i2.30
4. IETF. (2020). *RFC 7348 - Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks*. IETF. https://datatracker.ietf.org/doc/rfc7348/
5. Li, W., Liu, J., Wang, S., Zhang, T., Zou, S., Hu, J., Jiang, W., & Huang, J. (2021). Survey on Traffic Management in Data Center Network: From Link Layer to Application Layer. *IEEE Access*, *9*, 38427–38456. https://doi.org/10.1109/access.2021.3064008

6. Liu, G., Liu, Y., Meng, Q., Wang, B., Chen, K., & Shen, Z. (2025). Traffic load balancing in data center networks: A comprehensive survey. *Computer Science Review*, *57*, 100749. https://doi.org/10.1016/j.cosrev.2025.100749

7. RFC. (2018). *RFC 8365 - A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)*. RFC. https://www.rfc-editor.org/info/rfc8365

8. Zhang, Q., Zhao, G., Xu, H., Yu, Z., Xie, L., Zhao, Y., Qiao, C., Xiong, Y., & Huang, L. (2022). *Zeta: A scalable and robust East-West communication framework in Large-Scale clouds*. 1231–1248. https://www.usenix.org/conference/nsdi22/presentation/zhang-qianyu