Global Academic Journal of Economics and Business

Available online at https://www.gajrc.com
DOI: https://doi.org/10.36348/gajeb.2026.v08i04.038

OPEN ACCESS

ISSN: 2706-9001 (P)
ISSN: 2707-2584 (O)

Original Research Article

Resilient Network Segmentation Using EVPN-VXLAN and IP Fabric Architectures for Saudi Critical Data Centers under Vision 2030

Zakiuddin Mohammed[1]*
[1]Independent Researcher

***Corresponding Author**
Zakiuddin Mohammed
Independent Researcher

Article History
Received: 17.06.2026
Accepted: 11.08.2026
Published: 12.08.2026

Abstract: Saudi critical data centres are becoming strategic platforms for digital government, industrial automation, cloud services and data-intensive national programmes. Their network architecture must therefore sustain service continuity while constraining lateral movement, isolating failures and producing auditable evidence of policy enforcement. This review examines how Ethernet Virtual Private Network and Virtual Extensible LAN overlays, operating across routed leaf-spine IP fabrics, can provide a resilient segmentation foundation for that requirement. It synthesises protocol standards, security guidance and Saudi regulatory controls published between 2020 and 2025. The analysis distinguishes transport resilience from security isolation, showing that neither equal-cost forwarding nor a large overlay namespace is sufficient without explicit policy, controlled route propagation, authenticated infrastructure, service insertion and tested recovery procedures. It identifies four design principles: keep the underlay simple and failure-bounded; express tenant, application and trust boundaries in separate virtual routing domains; use control-plane learning and suppression mechanisms to reduce unnecessary flooding; and connect segmentation intent to operational telemetry, configuration assurance and incident response. A Saudi-oriented reference architecture is proposed for active-active critical sites, with distributed anycast gateways, multihoming, rapid designated-forwarder recovery, policy groups, controlled inter-site exchange and independent management-plane protection. The review concludes that EVPN-VXLAN should be treated as a programmable enforcement substrate rather than a security product. Its contribution to Vision 2030 depends on governance that links each segment, route and exception to data classification, criticality, recovery objectives and accountable ownership.

Keywords: EVPN, VXLAN, IP Fabric, Network Segmentation, Critical Data Centres, Cyber Resilience, Saudi Vision 2030, Zero Trust.

1. INTRODUCTION

Saudi Arabia's digital transformation has moved data centres from supporting infrastructure to nationally significant production environments. Digital government platforms, financial services, health systems, energy operations, logistics, artificial intelligence and industrial control increasingly depend on shared compute and cloud ecosystems. The Vision 2030 annual reporting framework links economic diversification to sustained investment in digital infrastructure, while the National Industrial Development and Logistics Program highlights data centres and advanced technologies as enablers of industrial growth [29, 30]. In this setting, a network outage or uncontrolled east-west compromise can affect more than one application. It can interrupt

public services, expose regulated data, disrupt operational technology dependencies and weaken confidence in national digital services. Architecture must therefore combine availability, containment, evidence and recoverability rather than optimising bandwidth alone.

Traditional data-centre segmentation relied on virtual local area networks, access lists and hierarchical switching domains. That model remains useful at small scale, but it becomes operationally fragile when thousands of workloads move between racks, availability zones and sites. Large Layer 2 fault domains amplify broadcast events and complicate deterministic recovery, while manual configuration creates drift between intended and actual policy. VXLAN increases the logical segmentation space by carrying Ethernet frames across routed infrastructure, and EVPN supplies a multiprotocol BGP control plane for endpoint, prefix and multihoming information. The result is a separation between a stable IP underlay and a service-oriented overlay that can support workload mobility, multi-tenancy and policy distribution [1-7]. This separation is particularly relevant where critical applications must be isolated without building a physically separate network for every trust domain.

Resilience, however, is not an automatic property of an overlay. A fabric may reconverge quickly yet still propagate an incorrect route, extend an unsafe broadcast domain or permit lateral movement between workloads with different criticality. Equally, aggressive security controls can create hidden single points of failure when all inter-segment traffic is forced through an undersized firewall cluster. The architectural problem is therefore socio-technical: protocols, topology, policy, automation, operational ownership and regulatory obligations must work as one system. Current standards increasingly address interconnect, operations and maintenance, integrated routing, prefix advertisement, multicast optimisation, mobility and designated-forwarder recovery [3-15]. Security guidance simultaneously emphasises explicit trust decisions, least privilege, authenticated management, asset visibility and continuous verification [16-20]. The Saudi controls add requirements for critical systems, cloud services, data lifecycle protection and operational technology [21-28].

This review evaluates how those strands can be integrated into a defensible architecture for Saudi critical data centres. It does not treat EVPN-VXLAN as a substitute for firewalls, identity services or endpoint controls. Instead, it asks how a standards-based fabric can make security zones more scalable, make failures more local and make recovery more predictable. The review also addresses a frequent design error: equating network virtualisation with microsegmentation. A virtual network identifier provides a forwarding boundary, but meaningful containment requires route-target discipline, policy at distributed gateways, group-based controls where supported, service insertion and deny-by-default communication between trust zones. The central argument is that resilient segmentation emerges when connectivity is deliberately constrained, observable and recoverable at every layer.

2. Aim, Objectives and Review Questions

The aim of this review is to develop a technically grounded and governance-aware model for resilient network segmentation in Saudi critical data centres using EVPN-VXLAN and routed IP fabric architectures. Four objectives guide the analysis. First, it clarifies the distinct resilience functions of the underlay, overlay, distributed gateway, multihoming and inter-site layers. Second, it evaluates how macrosegmentation and microsegmentation can limit attack propagation without creating unacceptable operational complexity. Third, it maps design choices to Saudi cybersecurity, cloud, data and critical-system obligations. Fourth, it proposes measurable implementation and assurance practices that connect architecture to day-two operations. The review addresses three questions: which architectural mechanisms most effectively contain faults and threats; how should segmentation be governed across active-active sites; and what operational evidence is required to demonstrate that intended isolation remains effective over time?

3. REVIEW METHODOLOGY

A structured integrative review was used because the topic spans protocol standards, data-centre engineering, cybersecurity architecture and national regulation. Searches were designed around five concept groups: EVPN and VXLAN; IP fabric and leaf-spine design; network segmentation and zero trust; data-centre resilience and operations; and Saudi critical infrastructure, cloud and data governance. Technical evidence was drawn from peer-reviewed work, Internet standards, national cybersecurity publications and authoritative implementation guidance issued between January 2020 and December 2025. The attached 2021 overview was used to establish baseline terminology and common deployment patterns, while later standards were prioritised for operational details such as interconnect, integrated routing, fast recovery and multihoming [1-15].

Material was included when it provided a defined mechanism, implementation requirement, security control or operational implication relevant to large-scale data-centre segmentation. Sources

were excluded when they were purely promotional, lacked a traceable publication date, repeated older descriptions without additional analysis, or focused on campus access without relevance to critical data-centre operations. Saudi materials were selected when they affected network architecture, cloud tenancy, critical systems, operational technology, data protection, service continuity or national digital infrastructure [21-30]. Evidence was coded into six themes: fault-domain design, control-plane behaviour, segmentation enforcement, inter-site resilience, management and observability, and regulatory alignment.

The synthesis compared mechanisms rather than products. For example, all-active multihoming was assessed as a redundancy pattern, not as a vendor feature; group-based policy was assessed as one method for identity-like segmentation, not as the only route to microsegmentation. Claims of benefit were retained only when the underlying mechanism was clear. The review did not invent performance measurements or infer universal convergence times because these depend on hardware, timers, scale, traffic mix and implementation quality. Instead, it derives design propositions that can be validated through laboratory testing, failure injection and operational telemetry. This approach supports a review paper that is analytically useful without presenting unperformed experiments as evidence.

4. Architectural Foundations of a Resilient Fabric

A resilient EVPN-VXLAN design begins with a deliberately uncomplicated Layer 3 underlay. In a leaf-spine topology, each leaf connects to every spine, creating multiple equal-cost paths and a predictable hop count. Routing failure is contained to links and nodes rather than being absorbed by a large spanning-tree domain. The underlay should carry infrastructure reachability only: loopbacks, point-to-point links, tunnel endpoint addresses and necessary management paths. Application subnets should remain in the overlay. This separation reduces the number of interacting states during incidents and makes faults easier to localise. It also permits independent scaling because additional leaf capacity can be introduced without redesigning application addressing.

Equal-cost multipath forwarding improves utilisation and removes dependence on a single aggregation path, but it must be paired with fast detection and stable routing policy. BFD for VXLAN provides a standards-based mechanism for rapid tunnel-path failure detection, while OAM requirements for EVPN define the need to verify service connectivity rather than merely physical reachability [2-4]. A critical site should monitor both planes: the underlay can be healthy while the overlay has stale endpoint state, and the overlay can be correct while a congested or asymmetric underlay degrades applications. Operational dashboards should therefore correlate BGP adjacency, tunnel reachability, route counts, MAC mobility, packet loss, latency and application synthetic tests.

The overlay uses virtual routing and forwarding instances to create macrosegmentation. Each tenant or trust zone receives a distinct Layer 3 context, and bridge domains are limited to subnets that genuinely require Layer 2 adjacency. Route distinguishers maintain uniqueness, while route targets govern import and export. This distinction is foundational: a route target is not merely a configuration label but a security-relevant policy object. An overly broad import policy can silently join environments that were intended to remain isolated. For critical workloads, route-target design should follow data classification and service dependency maps, with peer review for every cross-zone exchange. Automated validation should test that no prohibited route is imported and that only approved prefixes are advertised.

EVPN integrated routing and bridging enables distributed anycast gateways at leaf nodes, allowing local routing between subnets without hairpinning through a central core [5]. Local gateways reduce latency and remove a concentration point, but they also distribute enforcement responsibilities. The architecture must decide whether policy is applied at the source leaf, destination leaf, a service insertion point or a combination. Prefix advertisement procedures support routing of non-host prefixes and summarised services, helping separate endpoint mobility from broader reachability [6]. This is valuable for shared platforms because common services can be exported through controlled contracts instead of extending entire routing tables between zones.

Control-plane learning reduces reliance on flood-and-learn behaviour. EVPN advertises MAC and IP bindings, enabling remote tunnel endpoints to make forwarding decisions without first flooding unknown traffic. ARP and neighbour-discovery suppression further reduce broadcast pressure. Multicast proxy procedures can constrain membership signalling and improve scale for applications that genuinely require multicast [7]. These mechanisms enhance resilience by lowering background traffic and limiting the blast radius of endpoint discovery. They do not eliminate broadcast, unknown-unicast and multicast risk, however. Unknown-unicast flooding should be disabled or tightly bounded where applications do not require it, and storm controls should be tested under realistic failure conditions.

5. Segmentation as Threat and Failure Containment

Segmentation has three complementary levels. Macrosegmentation separates tenants, business services or criticality classes through virtual routing domains. Application segmentation separates tiers such as web, application, database and management services. Microsegmentation constrains communication between individual workload groups or identities, even when they share an IP subnet or routing instance. Table 1 distinguishes these layers. A defensible architecture uses all three selectively. Placing every workload in a unique segment may maximise theoretical isolation but create unmanageable policy. Conversely, broad environment-level segments leave excessive lateral paths. The practical unit of segmentation should be a stable security subject: an application tier, service function, administrative role or regulated data domain.

Zero-trust guidance requires access decisions to be explicit, least-privileged and continuously evaluated rather than inherited from network location [16-19]. EVPN-VXLAN contributes by supplying scalable forwarding boundaries and programmable attachment points. It does not provide user authentication or workload integrity by itself. Effective microsegmentation therefore combines network attributes with authoritative identity and asset context. Where group-based policy is available, security-group identifiers can travel with overlay traffic and drive distributed policy. Where it is not, equivalent outcomes can be achieved through host firewalls, distributed virtual switching, policy-aware firewalls or carefully managed access controls. The key requirement is consistency: the same logical relationship should produce the same decision regardless of workload location.

Service insertion remains necessary for traffic that requires stateful inspection, application-layer protection, malware analysis or data-loss controls. A common failure is to centralise every flow through a small security cluster, turning a segmentation improvement into an availability weakness. Service chains should therefore be capacity-tested, deployed redundantly and scoped to risk. East-west flows between high-criticality zones may warrant stateful inspection; low-risk flows within a tightly defined application tier may be adequately controlled by distributed policy. Route steering should fail closed for prohibited traffic, but continuity design must distinguish a secure fail-closed state from an uncontrolled outage. Bypass conditions, if permitted, need formal approval, time limits and complete logging.

All-active EVPN multihoming allows servers, appliances and service clusters to use multiple active attachments while preserving loop prevention and aliasing. This improves link utilisation and removes dependence on a single leaf. Yet multihoming introduces designated-forwarder behaviour for broadcast domains and requires careful split-horizon handling. Recent standards improve recovery and clarify redundancy modes, including fast designated-forwarder recovery, extended mobility procedures, selectable split-horizon signalling and port-active redundancy [12-15]. These developments are operationally important for critical sites because convergence must preserve both forwarding and isolation. A recovered path that temporarily duplicates traffic, loops broadcast frames or imports an unintended segment can be more damaging than a brief, bounded interruption.

Failure containment should be designed as a hierarchy. A host failure should affect one workload; a leaf failure should affect attached endpoints only; a spine failure should reduce path diversity without interrupting services; a border failure should preserve local operation; and a site failure should trigger application-level recovery without extending the fault to the surviving site. This hierarchy requires limits on Layer 2 extension. Stretching a bridge domain across sites may simplify a legacy cluster, but it couples broadcast, endpoint mobility and failure state across geography. The default should be routed inter-site connectivity with selective, justified Layer 2 extension. Every stretched segment should have an owner, a technical dependency, a documented removal plan and tested behaviour during partial partition.

Table 1: Segmentation layers and their resilience contribution

Layer	Primary boundary	Typical enforcement	Resilience contribution	Principal risk
Macrosegmentation	Tenant, criticality or data domain	VRF, route targets, firewall contracts	Contains routing faults and broad lateral movement	Route leakage through incorrect import/export policy
Application segmentation	Application tiers and shared services	Separate subnets, distributed gateway policy, service insertion	Limits compromise to a defined service dependency	Over-broad shared-service access or hidden tier coupling

Layer	Primary boundary	Typical enforcement	Resilience contribution	Principal risk
Microsegmentation	Workload groups or identities	Group policy, host controls, distributed firewalling	Restricts same-segment east-west movement	Policy sprawl, identity drift and troubleshooting delay
Management segmentation	Network and security administration	Dedicated VRF/OOB network, privileged access controls	Prevents production compromise from controlling the fabric	Shared credentials, exposed APIs or unmonitored break-glass access
Inter-site segmentation	Site, region and recovery domain	Border gateways, route filtering, selective DCI	Stops local faults and route churn spreading geographically	Uncontrolled Layer 2 stretch and dependency on common services

6. Inter-Site Architecture and Operational Continuity

Saudi critical data centres frequently require geographic diversity for service continuity, disaster recovery and data residency. EVPN overlay interconnect standards provide methods for exchanging reachability between fabrics while preserving administrative separation [3]. The preferred pattern is a multi-site architecture with independent underlay and overlay control planes inside each site, connected through redundant border gateways. Site-local faults should not cause uncontrolled route churn elsewhere. Border gateways should summarise or filter reachability, enforce route-target policy and prevent direct propagation of internal infrastructure state. This creates a deliberate failure boundary between sites rather than one geographically stretched fabric.

Active-active applications benefit from local ingress, local egress and deterministic path selection. Anycast service addresses and distributed gateways can provide locality, but symmetric stateful services may require flow-aware steering. The architecture should avoid assumptions that the network alone can make an application active-active. Databases, queues and stateful platforms need application-level replication semantics, conflict handling and consistency objectives. Network design supports those mechanisms by preserving reachability, quality and isolation. Recovery time and recovery point objectives must therefore be translated into testable network dependencies: route convergence, session persistence, load-balancer health, firewall state, DNS behaviour and storage replication paths.

Mobility is useful when it reduces maintenance disruption, but uncontrolled mobility can hide architectural debt. EVPN mobility procedures identify endpoint moves and sequence advertisements to suppress stale paths [12]. In a critical environment, frequent moves should be investigated because they may indicate orchestration instability, duplicate addressing or malicious activity. Thresholds for move counts and rapid oscillation should feed monitoring and incident workflows. The same telemetry can verify planned migration. A workload should not be considered successfully moved until its old attachment has withdrawn, its new policy has been applied, service health has passed and no unintended route remains in another zone.

Multicast and broadcast deserve explicit continuity planning. Some industrial, market-data and discovery systems depend on multicast, but default flooding across sites creates scale and security problems. EVPN multicast optimisation and proxy mechanisms allow membership-aware forwarding [7-11]. Designers should separate essential multicast groups by application and criticality, restrict source and receiver membership, and monitor unexpected joins. For non-multicast applications, suppression and routed boundaries should be preferred. This reduces the probability that a control-plane event becomes a data-plane storm during recovery.

7. Alignment with Saudi Critical-Infrastructure Obligations

Saudi controls frame segmentation as part of a wider accountability system. The Operational Technology Cybersecurity Controls require tailored protection for industrial control environments, while the Data Cybersecurity Controls address protection throughout the data lifecycle [21-22]. Critical data centres often host both conventional enterprise services and platforms that support operational systems. Those environments should not share unrestricted routing. Separate virtual routing domains, dedicated management paths and tightly controlled broker services can preserve necessary data exchange without allowing enterprise compromise to reach control networks. Where an application bridges both domains, its interfaces should be treated as high-risk conduits with protocol validation, monitoring and clear ownership.

The implementation guides for essential and critical-system controls emphasise context-specific application, asset understanding, secure configuration, monitoring, resilience and evidence [23, 24]. These expectations map directly to fabric governance. Every virtual network, routing instance, route target, policy group and inter-zone exception should be represented in an authoritative inventory. Configurations should be generated from approved intent where possible, checked before deployment and compared continuously with device state. Administrative access to route reflectors, controllers, border gateways and security-policy systems should use separate identity controls, multifactor authentication, privileged-session logging and restricted management networks. Compromise of the management plane can invalidate every data-plane boundary simultaneously.

The Essential Cybersecurity Controls and Cloud Cybersecurity Controls strengthen the case for layered segmentation, third-party governance, continuous compliance and data-localisation awareness [25, 26]. A cloud tenant using a shared facility may not control the provider underlay, but it remains responsible for defining its trust zones, monitoring exposure and validating provider commitments. Conversely, a national cloud or private-cloud operator must ensure that tenant isolation is enforced consistently across physical and virtual infrastructure. Route leakage between tenants should be treated as a high-severity security incident, not merely a configuration defect. Validation should include negative tests proving that prohibited communication fails.

Data governance standards and cloud regulations introduce further architectural consequences [27, 28]. Data classification should influence segment placement, encryption requirements, inspection paths and permitted destinations. Highly sensitive datasets may require dedicated virtual routing contexts, local key-management dependencies and restricted replication. Cloud service regulations also make contractual service levels, portability and provider responsibilities relevant to network design. A fabric that is technically resilient but opaque to customers cannot support effective assurance. Tenants need evidence about logical separation, incident notification, continuity testing and changes that affect connectivity or data location.

Vision 2030 does not prescribe a specific data-centre protocol. Its relevance lies in the scale and criticality of the digital economy that the network must support [29, 30]. Architecture should therefore be judged by whether it enables safe growth: rapid onboarding without weakening isolation, multi-site continuity without uncontrolled coupling, automation without loss of accountability and innovation without bypassing national controls. EVPN-VXLAN aligns with these needs when segmentation is governed as a service with measurable objectives. It becomes misaligned when overlay flexibility is used to create undocumented connectivity or when speed of provisioning takes precedence over risk acceptance.

8. Integrated Reference Architecture and Assurance Model

Figure 1 presents a reference architecture for two Saudi critical sites. Each site has a routed leaf-spine underlay with redundant spines, leaf pairs for compute, dedicated service leaves and border gateways. EVPN route reflectors are logically separated and deployed redundantly. Distributed anycast gateways terminate application subnets at the leaf. Virtual routing domains represent critical services, regulated data, shared platforms, management, security tooling and operational technology integration. Inter-zone communication is denied by default and enabled through reviewed contracts. Traffic requiring advanced inspection is steered through redundant service nodes. The sites exchange selected prefixes through border gateways rather than extending the whole control plane.

The architecture separates five planes. The forwarding plane carries application traffic. The control plane distributes validated reachability. The policy plane defines route import, security groups and service paths. The management plane configures and monitors infrastructure. The assurance plane independently evaluates whether deployed state matches approved intent. This separation matters because a controller reporting success is not sufficient evidence that every device enforced the change. Assurance should retrieve state from devices, flow telemetry, firewalls and synthetic probes. Discrepancies should block promotion or trigger rollback.

A change pipeline should begin with a machine-readable request containing business owner, application dependency, source group, destination service, protocol, data classification, expiry and rollback plan. Automated checks should detect overlapping address space, unsafe route-target imports, policy shadowing, missing logging and capacity risk. A staged deployment can then apply the change to a limited scope, execute positive and negative tests, and promote only after evidence is captured. Emergency changes need the same logging and retrospective review. Temporary connectivity must expire automatically because forgotten exceptions are a common source of accumulated lateral exposure.

Figure 2 illustrates a closed assurance loop: observe, detect, decide, contain, recover, verify and learn. Observation includes control-plane state, flow records, endpoint identity, configuration changes and service health. Detection correlates anomalies such as route spikes, unexpected mobility, policy denies, BFD transitions and new east-west paths. Decision logic classifies whether the event is a fault, attack, capacity issue or approved change. Containment can withdraw a route, quarantine a group, disable a tunnel, isolate a site edge or steer traffic through inspection. Recovery restores the minimum necessary service. Verification proves both availability and isolation before closure.

Table 2 maps architectural controls to operational evidence. The evidence model is essential because compliance cannot be inferred from design documents alone. A route-target matrix shows intended connectivity, but device route tables prove actual import. A segmentation policy shows permitted flows, but sampled flow telemetry and negative probes show enforcement. A redundancy diagram shows dual paths, but failure-injection results show recovery. Evidence should be timestamped, retained according to policy and linked to change records. This transforms the fabric from a collection of configurations into an auditable control system.

Key performance indicators should balance availability and containment. Availability measures include packet loss during single failures, convergence time, path diversity, control-plane stability and service latency. Containment measures include the number of reachable trust zones from a compromised workload, time to quarantine, prohibited-flow success rate, route-leak detection time and blast-radius size. Operational quality measures include configuration drift, failed change rate, expired exceptions, unresolved policy conflicts and percentage of segments with current owners. A design that improves uptime while expanding unnecessary reachability is not resilient; it has merely shifted risk from outage to compromise.

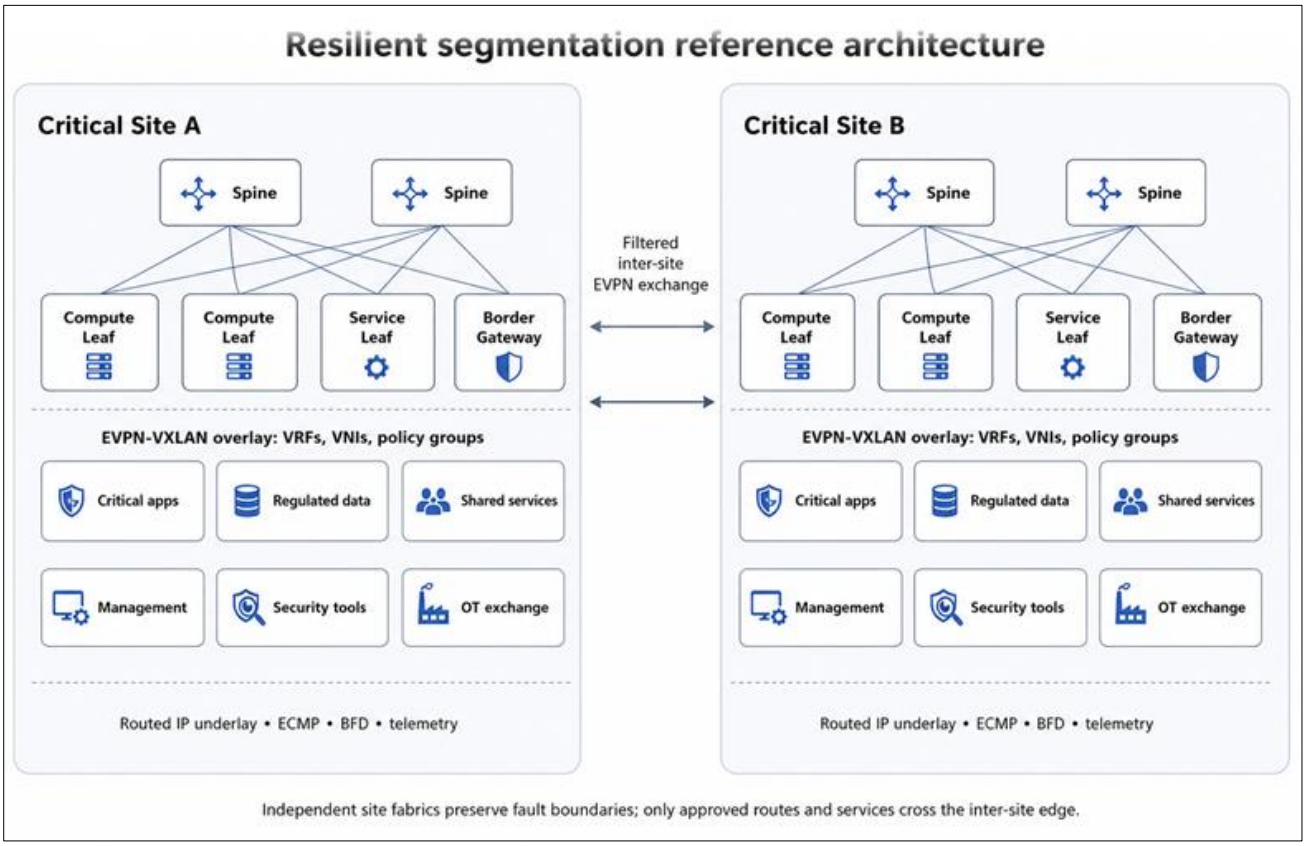

Figure 1: Saudi critical data-centre reference architecture using independent EVPN-VXLAN fabrics and controlled inter-site exchange

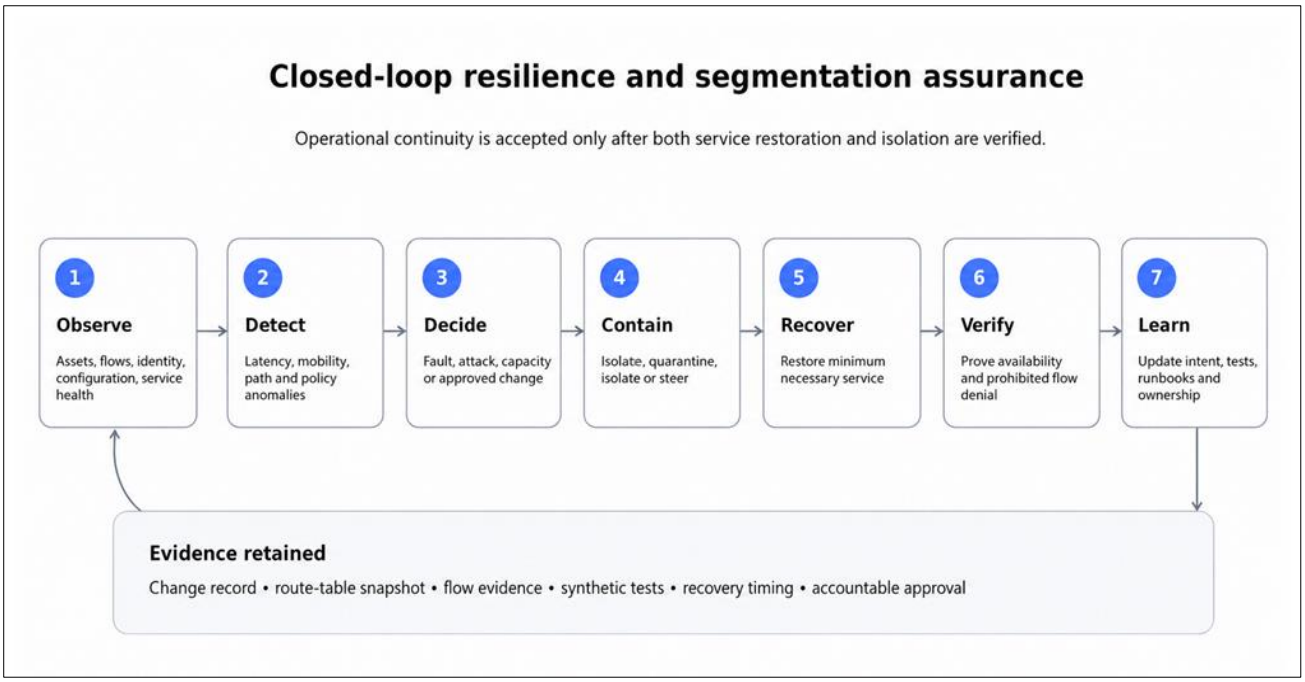

Figure 2: Closed-loop operational assurance for detection, containment, recovery and verified restoration

Table 2: Architecture-to-evidence mapping for resilient segmentation assurance

Control objective	Architecture mechanism	Operational evidence	Review trigger
Prevent cross-zone route leakage	Separate VRFs, constrained route targets, border filtering	Route-target matrix, device RIB/FIB snapshots, negative reachability tests	Any route-policy change, new shared service or tenant onboarding
Maintain path availability	Leaf-spine ECMP, redundant VTEPs, BFD, all-active multihoming	Failure-injection results, path telemetry, packet-loss and convergence records	Hardware replacement, timer change or capacity threshold breach
Contain lateral movement	Distributed policy, application tiers, service insertion, deny-by-default contracts	Flow logs, policy counters, denied-flow probes, exception register	New application release, identity-source change or incident finding
Protect management and control planes	Dedicated management VRF or out-of-band path, strong identity, restricted APIs	Privileged-session logs, AAA records, configuration signatures, controller audit trail	Administrator change, platform upgrade or credential event
Preserve site independence	Independent fabrics, redundant border gateways, selective inter-site advertisements	Partition tests, route-volume limits, recovery runbooks, site failover evidence	DR exercise, new DCI service or application recovery redesign
Demonstrate regulatory alignment	Asset ownership, data classification, mapped controls and retained evidence	Control mapping, approved design, test artefacts, remediation records	Annual assessment, major change or regulator/auditor request

9. DISCUSSION AND RESEARCH AGENDA

The evidence supports a clear distinction between fabric resilience and organisational resilience. Protocol mechanisms can provide multipath forwarding, rapid failure detection, endpoint mobility and controlled route distribution, but they cannot determine which applications should communicate or who may accept residual risk. Those decisions require service ownership, data classification and incident governance. The strongest architecture is therefore not the one with the largest number of virtual networks. It is the one in which every boundary has a purpose, every exception has an owner and every recovery action is tested.

Several trade-offs remain. Distributed gateways improve locality but expand the enforcement surface. Multi-site overlays improve continuity but risk coupling failures. Automated intent reduces manual error but can propagate a

flawed model at machine speed. Microsegmentation reduces lateral movement but can increase troubleshooting time and policy complexity. These trade-offs should be resolved through risk-based patterns rather than universal rules. High-criticality services justify stronger isolation, independent management and more frequent testing. Lower-risk shared services may use broader segments if monitoring and application controls remain strong.

Future research should measure EVPN-VXLAN security and recovery under realistic combined events rather than isolated device failures. Useful studies would examine route leakage during controller compromise, policy consistency during workload mobility, multicast containment during site partitions, and the interaction between BFD timers, control-plane load and application recovery. Saudi-specific research should also evaluate how data classification and critical-system controls can be translated into machine-readable network intent. A shared assurance vocabulary would help operators, auditors and regulators compare isolation, evidence and recovery without depending on vendor-specific terminology.

Implementation should distinguish configuration correctness from behavioural correctness. A fabric may contain the expected route targets and access rules yet behave differently when links flap, endpoints duplicate addresses, security appliances lose state, or an orchestration platform retries a failed transaction. Assurance programmes should therefore combine static analysis with controlled dynamic tests. Static checks confirm syntax, dependency and policy consistency; dynamic checks confirm forwarding, denial, recovery and observability under realistic conditions. Test cases should include unauthorised east-west attempts, route withdrawal, border isolation, leaf replacement, service-node failure and management-plane lockout. Results should be compared with agreed service objectives and retained as evidence. This practice is especially valuable before major national events or large application migrations, when simultaneous changes can increase shared risk. It also improves engineering learning because post-event reviews can compare predicted and observed behaviour, refine automation safeguards and remove obsolete exceptions. Architecture reviews should additionally examine dependency concentration. Shared route reflectors, identity stores, time sources, domain-name services, telemetry collectors and certificate authorities can become systemic failure points even when the forwarding fabric is physically redundant. Their placement, replication and recovery procedures should match the criticality of dependent services. Operators should define degraded modes for periods when a controller or analytics platform is unavailable. Devices must retain safe forwarding and policy state, while changes are suspended until trustworthy coordination returns. Training and runbooks should help network, security, cloud and application teams interpret events consistently. Shared operational language reduces delay, ambiguity and unsafe recovery decisions during incidents.

10. CONCLUSION

EVPN-VXLAN and routed IP fabrics provide a strong architectural basis for resilient segmentation in Saudi critical data centres. Their principal value is not simply a larger segment identifier or easier workload mobility. It is the ability to separate transport from service policy, distribute gateways, use all available paths, exchange reachability through a control plane and create deliberate failure boundaries. These capabilities can support Vision 2030 digital growth while aligning with national expectations for critical systems, cloud services, operational technology and data protection.

That value depends on disciplined implementation. The underlay should remain simple, observable and independently recoverable. Overlay routing domains should follow trust and data boundaries. Route targets, policy groups and inter-site advertisements should be treated as security controls. Layer 2 extension should be exceptional. Multihoming, fast recovery and service insertion should be tested under compound failures. Management and assurance planes should be isolated from production and capable of detecting drift. Finally, compliance should be demonstrated through operational evidence rather than assumed from diagrams.

A resilient network is one that continues the right services while preventing the wrong communications. For Saudi critical data centres, this requires availability and containment to be engineered together. EVPN-VXLAN can supply the programmable substrate, but governance determines whether that substrate becomes a controlled national capability or an increasingly complex source of hidden connectivity. The recommended approach is therefore architecture by explicit intent, deployment by verified automation and operation by continuous evidence.

REFERENCES

1. George, A. S., and George, A. S. H. (2021). A brief overview of VXLAN EVPN. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 9(7), 1-12. https://doi.org/10.17148/IJIREEICE.2021.9701

2. Internet Engineering Task Force. (2020). RFC 8971: Bidirectional Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN).
3. Internet Engineering Task Force. (2021). RFC 9014: Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks.
4. Internet Engineering Task Force. (2021). RFC 9062: Framework and Requirements for Ethernet VPN (EVPN) Operations, Administration, and Maintenance (OAM).
5. Internet Engineering Task Force. (2021). RFC 9135: Integrated Routing and Bridging in Ethernet VPN (EVPN).
6. Internet Engineering Task Force. (2021). RFC 9136: IP Prefix Advertisement in Ethernet VPN (EVPN).
7. Internet Engineering Task Force. (2022). RFC 9251: Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN).
8. Internet Engineering Task Force. (2022). RFC 9315: Intent-Based Networking - Concepts and Definitions.
9. Internet Engineering Task Force. (2023). RFC 9489: Label Switched Path Ping Mechanisms for EVPN and Provider Backbone Bridging EVPN.
10. Internet Engineering Task Force. (2024). RFC 9543: A Framework for Network Slices in Networks Built from IETF Technologies.
11. Internet Engineering Task Force. (2024). RFC 9625: EVPN Optimized Inter-Subnet Multicast Forwarding.
12. Internet Engineering Task Force. (2025). RFC 9721: Extended Mobility Procedures for Ethernet VPN Integrated Routing and Bridging.
13. Internet Engineering Task Force. (2025). RFC 9722: Fast Recovery for EVPN Designated Forwarder Election.
14. Internet Engineering Task Force. (2025). RFC 9746: BGP EVPN Multihoming Extensions for Split-Horizon Filtering.
15. Internet Engineering Task Force. (2025). RFC 9786: EVPN Port-Active Redundancy Mode.
16. National Institute of Standards and Technology. (2020). Special Publication 800-207: Zero Trust Architecture.
17. National Institute of Standards and Technology. (2022). Special Publication 800-215: Guide to a Secure Enterprise Network Landscape.
18. National Institute of Standards and Technology. (2024). The Cybersecurity Framework 2.0.
19. Cybersecurity and Infrastructure Security Agency. (2023). Zero Trust Maturity Model, Version 2.0.
20. National Security Agency. (2022). Network Infrastructure Security Guidance.
21. National Cybersecurity Authority. (2022). Operational Technology Cybersecurity Controls (OTCC-1:2022). Kingdom of Saudi Arabia.
22. National Cybersecurity Authority. (2022). Data Cybersecurity Controls (DCC-1:2022). Kingdom of Saudi Arabia.
23. National Cybersecurity Authority. (2023). Guide to Essential Cybersecurity Controls Implementation (GECC-1:2023). Kingdom of Saudi Arabia.
24. National Cybersecurity Authority. (2023). Guide to Cybersecurity Controls for Critical Systems Implementation (GCSCC-1:2023). Kingdom of Saudi Arabia.
25. National Cybersecurity Authority. (2024). Essential Cybersecurity Controls (ECC-2:2024). Kingdom of Saudi Arabia.
26. National Cybersecurity Authority. (2024). Cloud Cybersecurity Controls (CCC-2:2024). Kingdom of Saudi Arabia.
27. Saudi Data and AI Authority, National Data Management Office. (2021). Data Management and Personal Data Protection Standards. Kingdom of Saudi Arabia.
28. Communications, Space and Technology Commission. (2023). Cloud Computing Services Provisioning Regulations. Kingdom of Saudi Arabia.
29. Saudi Vision 2030. (2025). Vision 2030 Annual Report 2024. Kingdom of Saudi Arabia.
30. National Industrial Development and Logistics Program. (2024). Annual Report 2024. Kingdom of Saudi Arabia.

