Comparative Performance Analysis of Unified Carrier SDN Fabrics vs. Stitched SD-WAN Overlays for Distributed Data Centre Interconnects

Rainer Xavier Gomes
rainer.gomes@os3.nl

August 24, 2026, 43 pages

Academic supervisor:	Dr. M. (Marios) Avgeris, Informatics Institute (UvA) m.avgeris@uva.nl
Examiner:	Dr. C. (Chrysa) Papagianni, Informatics Institute (UvA) c.papagianni@uva.nl
Daily supervisor:	R. (Robert) Dijkerman, Head of Solutions, Enterprise Europe, NOKIA robert.dijkerman@nokia.com
External supervisor:	J. (Jaume) Figueras Solanilla, Customer Solutions Architect, Enterprise Europe, NOKIA jaume.figueras_solanilla@nokia.com
Host organisation:	Nokia Solutions & Networks B.V., Antareslaan 1, 2132 JE Hoofddorp The Netherlands

Universiteit van Amsterdam
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Master Security and Network Engineering
https://www.os3.nl

Abstract

Distributed data centre interconnects (DCI) increasingly carry latency-sensitive, failure-resilient workloads across heterogeneous wide-area bearers. Most deployments today stitch an EVPN-VXLAN data centre fabric to a separate WAN transport at a domain boundary, leaving failure recovery to proceed independently across two control planes. This thesis asks what changes when that boundary is removed at the WAN transport layer, replacing the stitching point with a unified Segment Routing over IPv6 (SRv6) micro-segment (uSID) fabric, while holding the data centre fabric constant as a controlled variable. The comparison is carried out in two phases on a Nokia virtualised testbed spanning three heterogeneous WAN bearers (fiber, 5G, and satellite). Phase 1 implements a stitched IS-IS and LDP transport with EVPN-VXLAN tenant isolation; Phase 2 replaces the WAN leg with a unified SRv6 uSID domain using Topology-Independent Loop-Free Alternate (TI-LFA) fast reroute and Flexible Algorithm (Flex-Algo) delay-metric path steering, together with a Path Computation Element (PCE) intended for centralised topology visibility. The two phases run sequentially on the same topology and traffic schedule, mirroring how an operator would validate an existing baseline before testing a candidate replacement. Failure recovery is the clearest difference between the two architectures: the unified fabric's pre-installed TI-LFA backup path restores traffic within milliseconds, against a multi-second reconvergence in the stitched baseline, a gap of more than three orders of magnitude even when measured against the worst observed case rather than the average. Steady-state latency and throughput show no clear advantage for either architecture. At this testbed's single-hop scale, the WAN-leg encapsulation overhead, measured directly from packet captures, is 32 bytes higher for the unified SRv6 uSID transport than for the legacy MPLS label stack, since uSID's compression advantage only emerges over multiple hops. A final finding concerns the maturity of the unified architecture itself: its distributed, per-node mechanisms, such as Flex-Algo path differentiation, are ready to deploy and verify today, whereas its centralised control plane was taken in this work as far as established, synchronised PCEP and BGP-LS sessions, with end-to-end PCE-driven path steering left as the next stage. This uneven maturity has direct implications for how operators should sequence a migration toward centralised control: the resilience gains are available immediately through distributed fast reroute, while the centralised path-computation layer can be adopted incrementally as it matures.

Contents

Chapter 1

Introduction

Over the past decade, the way enterprises and cloud providers build infrastructure has changed substantially. Workloads that once ran in a single data centre now run across multiple geographically distributed facilities, driven by requirements around fault tolerance, data residency, and the flexibility multi-cloud environments offer [1]. At the same time, the shift towards microservices-based application architectures means a single user-facing request can involve dozens of internal service calls spanning both intra- and inter-data-centre paths [2], and the growth of distributed AI training and inference workloads adds further pressure for consistent low-latency connectivity between accelerators [3]. The network tying these environments together is the Data Centre Interconnect (DCI), and its design choices directly shape which applications can run across distributed infrastructure. Latency-sensitive workloads need consistent round-trip times, data-heavy workloads need high goodput [4], and anything required to survive a failure needs fast recovery.

The DCI architectures most operators run today reflect the historical separation between data centre networking and wide-area networking: two disciplines that developed independently and were later integrated at a boundary, rather than being designed from the outset for the operational requirements that distributed workloads now impose. The result is a layered architecture: a DC segment running EVPN-VXLAN for multi-tenant isolation [5], connected to a WAN segment managed by a separate SD-WAN overlay controller. This boundary introduces three concrete problems. First, an encapsulation cost: traffic crossing from the DC into the WAN picks up a second header on top of VXLAN, a cost that compounds further where IPsec is also applied for WAN-link confidentiality [4]. Second, failure recovery is split across two independent control planes that converge separately with no shared topology state, so end-to-end recovery is bottlenecked by whichever domain is slower. Third, troubleshooting requires correlating telemetry from two separate management planes, neither of which has visibility into the other.

A distinction must be drawn here between two senses of "SD-WAN". Enterprise SD-WAN is a customer-premises-equipment (CPE) overlay model [1]: tunnels (typically IPsec or GRE over VXLAN) are built between branch and hub appliances over arbitrary, often third-party, transport, with a central controller selecting paths from application-level SLA probes. This model is well matched to connecting many branch sites to a few hubs over uncontrolled internet underlay, but it fits a distributed data centre environment poorly. It treats the WAN as an opaque tunnel mesh with no shared link-state view of the DC fabric, so it cannot participate in the EVPN-VXLAN control plane, cannot offer the sub-second, topology-aware fast reroute that a link-state IGP provides, and adds a further encapsulation layer on top of the VXLAN the DC fabric already imposes. For this reason, the relevant baseline for a carrier-grade distributed-DC interconnect is not an enterprise SD-WAN appliance but the carrier-stitched model: a link-state IGP (IS-IS) with label distribution (LDP) carrying an L3VPN service between the DC borders, which is the Phase 1 architecture evaluated in this thesis. The comparison this thesis draws is therefore between two carrier-grade data planes, the stitched IS-IS + LDP baseline and the unified SRv6 uSID fabric, rather than between a unified fabric and an enterprise CPE overlay; the latter is discussed in Chapter 2 as context for why it is not the appropriate baseline here.

A unified data plane that removes the separate stitching point at the WAN-transport boundary has the potential to address the first two of these problems directly. The third, telemetry correlation, is largely a function of which network operating systems sit at each layer rather than of how the WAN transport is encoded, and so falls outside the architectural change this thesis tests. The standardisation of Segment Routing over IPv6 (SRv6) [6], the micro-segment (uSID) compression extension [7], and the Path

Computation Element (PCE) architecture with Flexible Algorithm (Flex-Algo) support [8–10] together make WAN-transport unification technically feasible. In principle, SRv6 uSID can carry both DC tenant isolation and WAN path steering within a single 40-byte IPv6 header [11], though whether this reduces overhead in practice depends on path length, as measured directly from packet captures in Chapter 6; and a PCE with a topology database spanning both DC and WAN nodes can compute end-to-end paths and enforce differentiated forwarding policies, such as latency-optimised routing and multi-bearer path selection, without per-flow state in the core. This thesis tests how much of that potential is realisable on a specific multi-vendor platform combination rather than assuming it transfers directly into practice. This combination of WAN-transport unification components is referred to throughout as a **Unified Carrier SDN fabric**. The question being tested is whether it delivers measurable improvements over the stitched baseline in latency, throughput, and failure-recovery speed, and whether any such improvements are large enough to matter in practice.

This thesis approaches the problem in two experimental phases on a dedicated Nokia virtualised testbed. Phase 1 constructs a stitched baseline: a multi-bearer DCI topology running IS-IS + LDP with EVPN-VXLAN tenant isolation, representative of the gateway-based DC-WAN interconnect model in common deployment today [12]. Phase 2 replaces this with a unified SRv6 uSID carrier fabric, adding Topology-Independent Loop-Free Alternate (TI-LFA) fast reroute, Flex-Algo delay-metric path steering, and a PCE intended to provide centralised topology visibility; the extent to which that visibility was actually achieved is examined in Chapter 7. Both phases are evaluated on the same Nokia SR OS 25.7.R2 testbed under identical traffic and failure conditions, allowing a direct comparison of resilience, latency, and throughput characteristics.

1.1 Research Questions

The main research question guiding this thesis is:

> *To what extent can a Unified Carrier SDN architecture, in place of a traditional stitched SD-WAN overlay, improve the resilience of a Data Centre Interconnect (DCI), and what trade-offs does this introduce?*

This is supported by the following sub-questions:

- **RQ1.** *How can DC (EVPN-VXLAN) and WAN roles be mapped onto a unified SRv6 uSID fabric to unify the WAN-transport leg of a DCI, and what structural boundaries remain after this change?* Addressed in Chapter 5, where the Phase 2 topology and SRv6 uSID service configuration are described in detail. As discussed in Chapter 7, Section 7.4, the unification achieved here is specifically at the WAN-transport layer: the DC fabric is held constant as a controlled variable, and the DC-fabric-to-WAN handoff at the border PE remains a structural boundary in both phases. This question is answered with that scope in mind, not as a claim of end-to-end protocol unification.

- **RQ2.** *To what extent does a unified SRv6 uSID fabric affect failure-recovery behaviour, specifically in multi-path and multi-bearer scenarios, compared to traditional stitched architectures?* Addressed in Chapter 6, Sections 6.1.4 and 6.2.5, through failure-injection experiments on both Phase 1 and Phase 2 topologies. This question is scoped to the single fiber-bearer failure scenario evaluated in this thesis (Section 7.6), not to failure recovery in general.

- **RQ3.** *To what extent can a logically centralised Path Computation Element (PCE) using Flexible Algorithms (Flex-Algo) serve as a more deterministic alternative to decentralised path selection in a unified SRv6 fabric?* Addressed in Chapter 6, Section 6.2, and discussed further in Chapter 7, Section 7.4. Both PCEP and BGP-LS sessions between the OpenDaylight PCE and the border routers were brought to Established state, and Flex-Algo path differentiation is confirmed operational. This thesis scopes RQ3 to validating the centralised control plane the PCE depends on, together with the Flex-Algo result; end-to-end PCE-initiated SRv6 path instantiation builds on this and is set out as future work.

1.2 Contributions

This research was conducted in collaboration with Nokia Netherlands, whose industry supervision shaped the experimental scope and the choice of Nokia SR OS 25.7.R2 as the evaluation platform. The main

contributions of this thesis are:

1. A measurement of IS-IS + LDP reconvergence time on Nokia SR OS 25.7.R2, the same control-plane software image used on physical Nokia 7750 SR hardware, in a multi-bearer DCI testbed with heterogeneous WAN paths (fiber, 5G, and satellite). The measured 4.07 s reconvergence window, taken from IS-IS adjacency telemetry with the data-plane ICMP stream corroborating it, provides a stitched baseline against which SRv6 TI-LFA fast reroute is compared in Phase 2. To the author's knowledge, no prior controlled study has quantified this transition on Nokia SR OS in a complete virtualised DCI topology.
2. A direct convergence comparison between TI-LFA fast reroute [13] in the unified SRv6 fabric and IS-IS + LDP failover in the stitched overlay, triggered by the same backbone link failure and measured from the same application-level vantage point. Across 5 failure-injection trials at 1 ms probe resolution, TI-LFA bounds data-plane packet loss to a few milliseconds at most (mean 1.2 ms, maximum 3 ms), a conservative improvement of more than 1355× over the Phase 1 baseline.
3. An evaluation of Flex-Algo path steering [10, 14] and centralised PCE deployment in a multi-bearer scenario. Flex-Algo delay-metric path differentiation is confirmed operational, and the OpenDaylight PCE is brought up to established, synchronised PCEP and BGP-LS sessions with the border routers. The evaluation is scoped to establishing and validating this centralised control plane; end-to-end PCE-initiated SRv6 path instantiation builds on it and is set out as future work.
4. A fully documented Containerlab testbed [15] implementing both architectures as version-controlled declarative topology definitions, deployable from a single command on any host running Nokia SR-SIM,[1] together with two implementation observations encountered during the build. The first is general and vendor-independent: a minimum of two leaf nodes per data centre is required to genuinely exercise EVPN-VXLAN bridging, since with a single VTEP there is no second bridging endpoint for an EVPN Type-2 route to traverse. This is a property of EVPN itself and applies to any conformant implementation; it is detailed in Chapter 5. The second concerns address-family scope: the unified SRv6 path was evaluated with the IPv4 tenant service (End.uDT4) only, and the Phase 2 fabric was correspondingly addressed for IPv4 only, whereas the Phase 1 stitched transport carried a dual-stack 6VPE service. Extending the unified-path evaluation to a cross-DC IPv6 tenant service was kept out of scope for this iteration, is documented in Chapter 7, Section 7.5, and is identified as future work rather than as a limitation of either architecture.

All timing and throughput figures throughout this thesis are measured on the SR-SIM virtualised data plane, which forwards in software. They are valid as relative comparisons between the two architectures under identical conditions, not as representative of the absolute performance of physical Nokia 7750 SR hardware.

1.3 Outline

Chapter 2 reviews the relevant literature and identifies the gaps this thesis addresses. Chapter 3 covers the technical background needed to follow the experimental design. Chapter 4 establishes the research methodology. Chapter 5 describes the testbed, the two architectural configurations, and the measurement procedures. Chapter 6 presents the experimental results. Chapter 7 interprets the findings against the research questions and discusses limitations. Chapter 8 summarises the key findings.

[1] The complete testbed, Containerlab topology definitions, node configurations, traffic-generation scripts, and measurement pipelines are available at `https://github.com/rainx-10/dci-unified-srfabric`.

Chapter 2

Related Work

This chapter reviews the work most relevant to the problem this thesis addresses. It is organised around three threads: the performance and limitations of SD-WAN-based DCI architectures, the use of SRv6 and unified data planes for WAN transport, and the convergence and path-steering properties of SR-based networks. The chapter closes with a gap analysis.

2.1 SD-WAN Architectures and DC-WAN Integration

The standardised basis for the stitched DCI model evaluated in this thesis is RFC 9014 [12], which defines how network virtualisation overlay (NVO) networks running EVPN can be connected to a WAN, via either an integrated gateway (NVO and WAN edge functions on the same device) or a decoupled one (the two functions on separate systems), and specifies how EVPN routes are processed at the gateway boundary between the EVPN-Overlay and WAN domains. The integrated variant corresponds directly to the border PE node used in Phase 1 of this thesis. Among the gateway optimisations it defines, RFC 9014 introduces the Unknown MAC Route (UMR) specifically to contain MAC-table scale on the data centre network virtualisation edge, illustrating the kind of boundary-management cost that the stitched model carries and that a unified WAN-transport data plane sets out to reduce.

Troia et al. provide a recent survey of SD-WAN architecture [1], covering CPE design, underlay technologies, overlay encapsulation protocols (VXLAN, GRE, IPsec), traffic engineering, and network monitoring. They acknowledge that VXLAN adds per-packet overhead that can create bandwidth inefficiencies in constrained environments, identify the SRv6 header overhead from long IPv6 addresses as a scalability concern, and note that BFD can be configured with detection intervals as low as 50 ms. What the survey does not do is quantify either the per-packet encapsulation cost or the failure-recovery latency under controlled conditions, nor does it address unifying the WAN transport layer specifically while leaving the data centre fabric untouched. That gap is part of what this thesis addresses.

Which class of SD-WAN is the relevant comparison for a distributed data centre interconnect needs to be stated precisely, since the term covers two quite different deployment models. The enterprise SD-WAN model surveyed above is a customer-premises-equipment overlay: appliances at each site build encrypted tunnels (commonly IPsec or GRE over VXLAN) across arbitrary underlay transport, and a central controller selects paths using application-level SLA probes. This is effective for connecting many branch offices to a small number of hubs over uncontrolled, often third-party, internet underlay. In a distributed data centre environment running EVPN-VXLAN locally, however, the enterprise model fits poorly: the SD-WAN overlay has no link-state view of the DC fabric and cannot participate in its EVPN control plane; it relies on probe-driven path selection rather than the topology-aware sub-second fast reroute a link-state IGP provides, and it stacks a further tunnel encapsulation on top of the VXLAN the DC fabric already imposes. The architecture a carrier actually deploys to interconnect distributed data centres is therefore not an enterprise SD-WAN appliance but a stitched carrier model: a link-state IGP (IS-IS) with label distribution (LDP) carrying an L3VPN service between DC border PEs, which is the Phase 1 baseline in this thesis. RFC 9014, discussed above, standardises exactly this gateway-based interconnect. The comparison drawn in this thesis is accordingly between two carrier-grade WAN-transport data planes rather than between a unified fabric and an enterprise CPE overlay, and the limitations of the enterprise model in this setting are the reason it is not used as the baseline.

Akharchaf and Gao compare MPLS, SD-WAN, and SRv6 under AI-enhanced traffic engineering on a 24-node emulated topology [4], framing SD-WAN as carrying inherent overlay overhead with no native

QoS guarantees and noting that SRv6 needs larger per-packet headers and platform upgrades. Without AI optimisation, their simulated end-to-end latency for real-time traffic is 34.8 ms for MPLS, 41.2 ms for SRv6, and 52.5 ms for SD-WAN. These results come from Mininet with synthetic traffic classes; the study does not evaluate a complete DCI scenario with an actual DC fabric and WAN segment, nor does it measure link-state convergence under failure. This thesis addresses both gaps using a Nokia SR OS virtualised testbed.

Sezer et al. identify controller scalability as a central limitation of centralised SDN control, noting that exchanging network state between distributed nodes and a single controller adds latency that grows with network size, and that centralised controllers are an attractive target for denial-of-service and hijacking precisely because of that role [16]. This is directly relevant to the PCE centralisation trade-off evaluated in Phase 2, where an OpenDaylight PCE was deployed as a logically centralised path computation element for the unified WAN transport, with the extent of integration actually achieved described in Chapter 7, Section 7.4. Large-scale SD-WAN deployments such as Google's B4 [17] and Microsoft's SWAN [18] showed that centralised traffic engineering is viable at production scale by *decoupling* control-plane software from switch hardware, attributing their gains to global, coordinated path assignment from a centralised topology view rather than to any specific data-plane encapsulation. This motivates the centralised-visibility component of the PCE-based architecture evaluated here, independent of the SRv6 encapsulation choice itself.

Francois et al. established the practical foundations for sub-second IGP convergence in large IP networks through router-level measurement and large-scale simulation [19]. They decompose convergence time into its contributing factors, including failure detection, link-state packet origination, flooding, SPF computation, and RIB/FIB update, and identify the FIB update as the dominant cost at the router level, since it scales with the number of affected prefixes while detection and SPF computation stay comparatively fast. Their work predates SR and is scoped to the IGP itself, with no treatment of LDP label resignalling; it provides the analytical framework used to interpret the IS-IS-specific portion of the Phase 1 convergence measurement in Chapter 6, though the LDP component falls outside their scope.

2.2 SRv6 and Unified WAN Transport

RFC 8986 [6] defines the SRv6 network programming model and the endpoint behaviour library used throughout this thesis. The End.DT4 behaviour (realised as End.uDT4 in Phase 2 for VPRN tenant traffic) allows SRv6 to terminate L3VPN service without a separate encapsulation layer at the WAN boundary; the companion End.DT6 behaviour provides the equivalent IPv6 function but was not exercised in this iteration (Section 7.5). The foundational segment routing architecture, SID semantics, segment list processing, and the SR policy framework used by the PCE are defined in RFC 8402 [20].

The header overhead problem of standard SRv6 is directly addressed by Tulumello et al., who introduce the Micro SID (uSID) compression mechanism [7]. In standard SRv6, the full encapsulation cost for a segment list of length SL is $40+8+(SL-1)\times 16$ bytes (outer IPv6 header, fixed SRH portion, and $SL-1$ additional 16-byte segments, since the first segment is already in the destination address); uSID instead packs multiple short (2-byte) SID fragments into a single 16-byte container carried in the IPv6 destination address, so that a path of several waypoints can be encoded in a flat 40-byte outer header with no SRH (the exact number per container depends on the block length, as detailed in Chapter 3). Tulumello et al. report an encapsulation-size saving of approximately 58% relative to standard SRv6 already at a four-node segment list, rising towards roughly 70% as the number of segments increases; the advantage is specifically a multi-hop property, and the two schemes are arithmetically identical at SL=1. Phase 2 of this thesis uses the Nokia SR OS implementation of the same compression principle; the single-hop WAN scale evaluated here is precisely the case where uSID offers no compression advantage over an MPLS label stack, which the packet-capture measurement in Chapter 6 confirms directly.

Ventre et al. survey SR research covering standardisation, implementation across VPP, Linux, P4, and Nokia SR OS, and performance evaluation [21], establishing that SR-MPLS and SRv6 are production-ready across platforms, though without evaluating DCI convergence under multi-bearer failover. Abdelsalam et al. present SRPerf, an SRv6 performance evaluation framework quantifying forwarding throughput (No-Drop-Rate / Partial-Drop-Rate) on Linux and VPP, including fixing a genuine implementation bug in two Linux-kernel endpoint behaviours [22], a methodological reference for the measurement approach in Chapter 5. Scarpitta et al. demonstrate STAMP-based delay monitoring for SRv6 SD-WANs, showing that an eBPF implementation has a negligible impact on the forwarding throughput of a Linux software router [23]; their protocol, like STAMP generally, assumes synchronised clocks between sender and reflector rather than eliminating that requirement, and their measurement-overhead finding mo-

tivates preferring router-native counters over in-band active probing in the Phase 2 telemetry design (Chapter 4).

2.3 Fast Reroute, Convergence, and Path Computation

Topology Independent Loop-Free Alternate (TI-LFA) is the standard SR-native Fast Reroute mechanism [21]. TI-LFA pre-computes a backup path at the Point of Local Repair (PLR) that is guaranteed to be loop-free regardless of topology, using the post-convergence path as the repair path, which eliminates the double-reroute problem of earlier LFA variants. Because the backup is pre-installed in the forwarding plane, no control-plane processing is required on the recovery-critical path, which is the architectural reason TI-LFA can, in principle, restore traffic far faster than IGP reconvergence.

Singh et al. present a configuration and functional-verification study of TI-LFA with segment routing on Cisco IOS XRv routers [24], demonstrating correct label-stack behaviour across zero-, single-, and double-segment repair scenarios and SRLG protection via forwarding-table inspection; their work confirms TI-LFA's correctness but reports no recovery-time measurements, and the sub-50 ms figure in their introduction is cited from a separate survey as background, not as their own result. Roelens et al. evaluate TI-LFA's efficiency via discrete-event simulation on a 10-node, 14-link topology, measuring success rate, path stretch, and node repetitions rather than recovery latency [25]; their central finding is that TI-LFA backup paths can revisit nodes already traversed by the primary path under certain routing-policy and load combinations, increasing congestion without improving coverage. Neither paper measures recovery time against a stitched overlay baseline, which is precisely what this thesis measures in Chapter 6.

For path computation, RFC 4655 [8] defines the PCE architecture, in which a logically centralised element computes paths on behalf of head-end routers using a Traffic Engineering Database (TED) populated from IGP TE extensions and BGP-LS. The stateful PCE extensions in RFC 8231 [9] extend this model to allow the PCE to maintain an active LSP state database and to initiate or update SR policies via PCEP, which is the model used by the OpenDaylight PCE in Phase 2.

RFC 9350 [10] standardises Flexible Algorithms (Flex-Algo) for link-state IGPs. A Flex-Algo definition specifies a calculation type, metric type, and a set of administrative constraints distributed to all participating nodes via IGP Flex-Algo Definition (FAD) TLVs. Each node independently computes a Flex-Algo-specific forwarding plane, and a PCE can instantiate SR policies that steer traffic onto Flex-Algo-constrained paths without per-flow state in the core. RFC 8570 [14] defines the IS-IS TE metric extensions that advertise unidirectional link delay as a sub-TLV, which is the delay metric consumed by Flex-Algo 128 in Phase 2 to compute minimum-latency paths.

2.4 Research Gap

Putting this together, the existing literature tells a reasonably clear story. The gateway-based DCI stitching model adds encapsulation overhead and requires independent failure recovery in each domain. SD-WAN encapsulation overhead is widely acknowledged, and BFD gives sub-second detection, but neither overhead nor recovery latency has been quantified experimentally in a complete end-to-end DCI scenario under realistic DCN traffic. SRv6 uSID can compress the WAN-transport header significantly relative to standard SRv6 over multi-hop paths, though this advantage does not extend to single-hop paths, where the two schemes are equal. TI-LFA provides pre-computed, loop-free fast reroute with guaranteed topological coverage. Flex-Algo and stateful PCE together provide a standards-based framework for multi-objective path steering, with the maturity of the centralised side varying across multi-vendor combinations in practice.

What is missing is a controlled, empirical comparison that brings these threads together in a single experiment on a fixed data centre fabric. To the best of the author's knowledge, no published study has: (1) measured the latency and throughput differential between a stitched IS-IS + LDP WAN transport and a unified SRv6 uSID WAN transport under identical traffic loads while holding the data centre fabric constant; (2) directly compared TI-LFA recovery time against IS-IS + LDP failover under the same backbone link failure measured from the same application-level traffic stream; or (3) attempted PCE deployment alongside Flex-Algo path differentiation in a multi-bearer DCI context, establishing how far a centralised PCE can be brought up and integrated on current multi-vendor software combinations.

These gaps matter for three reasons. Operators considering a migration from a stitched WAN overlay to a unified SRv6 fabric need quantitative evidence of the resilience improvement, not just theoretical

arguments, since a recovery-time gap of the magnitude measured here changes the risk calculus for latency-sensitive workloads such as storage replication and distributed AI training. The encapsulation comparison grounds the architectural choice in concrete packet economics, including the finding that uSID's advantage is scale-dependent rather than universal. Establishing how far Flex-Algo and a centralised PCE can be deployed and integrated on Nokia SR OS gives reproducible evidence of how mature each part of the standards-based architecture currently is, useful to both the research community and to operators planning a migration.

This thesis addresses the first two gaps directly and approaches the third through what could be tested: Flex-Algo path differentiation is confirmed operational, and the centralised PCE is brought up to working, synchronised PCEP and BGP-LS sessions, establishing how far the centralised control plane can be taken on this multi-vendor combination. The headline result, developed in Chapter 6, is that the convergence improvement is substantial and measured rather than assumed: TI-LFA bounds the data-plane disruption window to a few milliseconds against a multi-second IS-IS + LDP baseline. The encapsulation comparison in the same chapter shows the opposite trade-off, with uSID's per-packet overhead measured at 32 bytes above the MPLS label stack at this testbed's single-hop scale, consistent with uSID's compression advantage being a multi-hop property that a compact topology cannot exhibit.

Chapter 3

Background

This chapter covers the technical foundations needed to follow the experimental design and results in later chapters. It begins with the DC fabric technologies that form the stitched overlay baseline (Section 3.1), then covers the WAN overlay and domain boundary model (Section 3.2), followed by the SRv6 unified data plane and its uSID compression extension (Section 3.3), the IS-IS path computation and Flexible Algorithm framework (Section 3.4), TI-LFA fast reroute (Section 3.5), and the emulation and measurement tools used throughout the testbed (Section 3.6). Concepts are introduced in roughly the order they first appear in the implementation description.

3.1 Data Centre Fabric: EVPN and VXLAN

Ethernet VPN (EVPN), standardised in RFC 7432 [5], is a BGP address family that distributes MAC and IP reachability information across a network using a common control plane. EVPN moves MAC learning into the control plane rather than the data plane, which eliminates broadcast-domain flooding across multi-site fabrics. Route Distinguishers (RDs) make routes from different tenants globally unique in BGP, while Route Targets (RTs) govern per-VRF import and export, providing multi-tenant isolation across the shared BGP infrastructure.

Two EVPN route types are relevant to the DC-fabric side of this thesis. Type-2 (MAC/IP Advertisement) routes advertise per-host MAC/IP bindings from each VTEP, enabling control-plane MAC learning and suppressing data-plane flood-and-learn across the fabric; this is the route type underlying the proxy-ARP/proxy-ND advertisement chain described in Chapter 5. Type-3 (Inclusive Multicast Ethernet Tag) routes advertise BUM group membership, letting VTEPs build multicast trees entirely in the control plane. Inter-DC host-subnet reachability, by contrast, is not carried by EVPN at all in either phase of this testbed: it is distributed via the BGP-IPVPN VPN-IPv4/VPN-IPv6 address families [26] between the border PE loopbacks, as described in Section 3.2.

Virtual eXtensible LAN (VXLAN), defined in RFC 7348 [27], provides the data-plane encapsulation for EVPN by extending Layer 2 segments over a Layer 3 IP underlay. The per-packet overhead totals 50 bytes over IPv4 or 70 bytes over IPv6, and the 24-bit VNI supports up to 16 million logical segments. This VXLAN encapsulation is part of the data centre fabric and is held identical in both experimental phases of this thesis (Chapter 4); the encapsulation comparison this thesis performs is at the separate WAN-transport stitching point, where Phase 2 replaces an MPLS label stack with a single SRv6 uSID header rather than touching the VXLAN layer itself.

3.2 WAN Overlay and Domain Stitching

SD-WAN builds a virtualised WAN overlay over one or more underlay transports by creating tunnels between gateway nodes at each site [1]. A centralised controller handles path selection and policy, typically using BFD [28] as the liveness mechanism and SLA probes to measure per-path delay and loss.

RFC 9014 [12] defines the DC-WAN gateway model that forms the conceptual basis for Phase 1 of this thesis. In this model, a DC-GW node sits at the boundary between the DC fabric and the WAN, terminating the DC-side EVPN-VXLAN encapsulation and re-originating tenant routes as WAN VPN prefixes. The two domains maintain independent control planes with no shared topology state, and recovery after a WAN link failure depends on the WAN's own liveness mechanism rather than on a

unified link-state IGP.

The Phase 1 implementation uses IS-IS and LDP as the WAN transport rather than a controller-driven SD-WAN overlay. The DC fabric runs BGP mac-vrf with IRB, and the two domains are stitched together at the border via a VPRN L3VPN service [26]. Path selection in the WAN is governed by IS-IS link metrics configured manually on the border nodes. This represents a traditional carrier-grade stitched architecture: the DC and WAN are independently managed, and failover requires IS-IS convergence and forwarding table updates before traffic resumes. The dual-stack (IPv4/IPv6) extension to this Phase 1 service, implemented using the BGP-MPLS IPv6 VPN extension (6VPE) [29] with the Extended Next-Hop Encoding capability [30] (advertising IPv6 VPN routes with an IPv4 next-hop), is described in Chapter 5.

3.3 Segment Routing over IPv6 (SRv6)

Segment Routing [20] is a source-routing architecture where the head-end node encodes the complete forwarding path for a packet as an ordered list of Segment Identifiers (SIDs). Transit nodes forward based on the active SID without maintaining per-flow state, which eliminates the distributed RSVP-TE state required by MPLS traffic engineering. The SRv6 variant [6] encodes SIDs as 128-bit IPv6 addresses and carries them in the IPv6 Segment Routing Header (SRH) [31].

Each SRv6 SID is a 128-bit IPv6 address split into a *Locator*, a topologically routable prefix that routes the packet to the correct node, and a *Function*, the instruction executed on arrival. The End function performs a next-hop lookup, End.DT4 performs a per-VRF IPv4 table lookup for L3VPN service delivery, and uN is the micro-segment-specific node function used in Phase 2. Transit P-routers that are not SID endpoints forward purely on the outer IPv6 destination address and never inspect the SRH. The per-packet overhead of standard SRv6 grows with segment-list length SL: the full encapsulation cost is $40 + 8 + (SL - 1) \times 16$ bytes, the outer IPv6 header, the fixed SRH portion, and $SL - 1$ additional 16-byte segments, since the first segment is already carried in the IPv6 destination address rather than duplicated in the segment list. This overhead can exceed VXLAN's at sufficiently long path lengths, which is the motivation for the uSID compression discussed next.

The Micro SID (uSID) mechanism [7] addresses this overhead problem by packing several 16-bit micro-instructions into a single 128-bit SID container, so that a multi-hop path can be expressed in one IPv6 address rather than a stack of full 128-bit SIDs. The number of micro-SIDs a container holds depends on the block length: with the 48-bit micro-segment block used in this thesis, the remaining 80 bits carry up to five 16-bit micro-SIDs per container (a 32-bit block, as in the widely cited F3216 format, carries up to six). Each micro-SID identifies a node or function within the block, and transit nodes shift the container left by 16 bits at each hop. Under uSID, encapsulation overhead becomes flat at 40 bytes (a bare IPv6 header, with the SRH itself omitted under reduced encapsulation) regardless of path length, rather than growing per segment as in standard SRv6. The resulting saving is a multi-hop property, quantified in Section 2.2; the two schemes are arithmetically identical at SL=1. Nokia SR OS implements uSID node-level forwarding as the `uN` endpoint behaviour, adjacency-level steering as `uA`, and per-VRF IPv4 table lookup as `uDT4`, all of which are used in Phase 2 as described in Chapter 5.

3.4 IS-IS, Traffic Engineering, and Flexible Algorithms

The control plane for the unified SRv6 fabric in Phase 2 combines three components: IS-IS as the link-state IGP for dynamic route distribution, BGP-LS for topology export to a centralised PCE, and Flexible Algorithms for associating alternative metric types with specific forwarding planes.

IS-IS [32] maintains a link-state database (LSDB) by flooding Link-State PDUs (LSPs) to all routers in the area, from which each router independently computes shortest paths using Dijkstra's SPF algorithm. When a link fails, the adjacent router generates an updated LSP, floods it to all peers, and each router independently re-runs SPF to update its forwarding table. The TE extensions of RFC 8570 [14] add sub-TLVs to IS-IS LSPs that advertise unidirectional link delay as a metric attribute, which Flex-Algo path computation consumes. BGP Link-State (BGP-LS), defined in RFC 9552 [33], enables IS-IS to export its LSDB to a BGP peer, giving the PCE a consistent global topology view, provided that the session actually establishes; whether it did so in this testbed is addressed in Chapter 5.

The Path Computation Element (PCE) architecture [8] defines a logically centralised element that computes paths on behalf of head-end routers using a Traffic Engineering Database populated from IGP TE extensions and BGP-LS. The stateful PCE extensions [9] allow the PCE to maintain active LSP

state and to initiate or modify SR policies via PCEP.

Flexible Algorithms (Flex-Algo), standardised in RFC 9350 [10], define per-algorithm SPF variants that use alternative link metric types. A Flexible Algorithm Definition (FAD) specifies a metric type, such as TE delay or IGP metric, a calculation type, such as shortest path or widest path, and optional administrative group constraints for link inclusion or exclusion. Each participating node distributes the FAD via IS-IS TLVs and independently computes a Flex-Algo-specific forwarding topology. In Phase 2, Flex-Algo 128 with delay metric is used to compute the minimum-latency path across the WAN, replacing manual metric configuration with automatic delay-aware path selection.

3.5 Fast Reroute: TI-LFA

Topology-Independent Loop-Free Alternate (TI-LFA) fast reroute [13] pre-computes a backup path at each Point of Local Repair (PLR) and installs it in the forwarding plane before any failure occurs. When the IS-IS module detects an adjacency loss, the pre-installed backup path is activated without waiting for IS-IS re-convergence.

TI-LFA has four properties that make it suitable for DCI fast reroute. First, the repair path is computed during topology initialisation rather than at failure time, so no control-plane processing is on the recovery-critical path. Second, TI-LFA uses the post-convergence optimal path as the repair path, so traffic routed via the backup does not need a second reroute after IS-IS converges, eliminating the double-reroute problem of earlier LFA variants. Third, unlike classic IP-FRR, TI-LFA provides guaranteed backup coverage in any two-connected network by using SR segment encoding to express an explicit repair path bypassing the failed element, even when no direct loop-free alternate exists. Fourth, the backup path is encoded as an SR segment list at the PLR, so no RSVP-TE tunnel state is required anywhere else in the network. These properties make very fast activation *possible* in principle; how fast it actually is on this specific testbed is an empirical question this thesis answers directly in Chapter 6 (mean 1.2 ms, maximum 3 ms across 5 trials), rather than one this chapter asserts a number for in advance, and is compared against the ≈4.07 s recovery time measured in Phase 1 at 200 ms probe resolution.

3.6 Testbed Tools and Platforms

Containerlab [15] orchestrates the testbed by creating a Docker container per node from a declarative YAML topology file, with reproducibility guaranteed since the topology and configuration files are version-controlled. Nokia SR Linux runs on the DC leaf and spine nodes (IXR-D2L model), and Nokia SR OS runs on the border PE and P-router nodes via the SR-SIM containerised simulator [34]. SR-SIM executes the same SR OS software image as physical Nokia 7750 SR hardware and supports IS-IS, BGP, LDP, SR-MPLS, SRv6 uSID, PCEP, and Flex-Algo. Its data-plane forwarding is software-based and throughput-limited, so all timing and throughput figures in this thesis are relative comparisons between architectures under identical conditions, not measurements representative of physical hardware. All programmatic interaction with SR OS nodes is handled via netmiko, which manages SSH sessions, the SR OS login banner, and MD-CLI mode switching; this proprietary I/O stack also means Linux `tc netem` does not affect SR OS containers, since it bypasses the kernel network layer entirely. The operational details of how each platform is interrogated and how failures are injected are described in Chapter 5, where they matter for reproducing the measurements.

TrafPy [35] generates the DCN bimodal traffic schedule used throughout this thesis, producing a reproducible mix of elephant and mouse flows from measured DC workload distributions with a configurable seed for exact repeatability across runs. The gNMIc collector, Prometheus, and Grafana form the telemetry stack for Nokia SR Linux nodes, which expose IS-IS and interface state via OpenConfig YANG paths [36]. Nokia SR OS nodes do not implement OpenConfig YANG models for IS-IS or MPLS state, so SR OS telemetry is instead collected via netmiko polling; this split is a property of the network operating systems deployed in this testbed and applies equally in both phases, independent of which WAN-transport architecture is in use.

Chapter 4

Methodology

This chapter describes the research methodology used to evaluate the comparative performance of a unified carrier SDN fabric against a stitched SD-WAN overlay for distributed DCI. It covers the experimental design, the choice of evaluation metrics, and the traffic generation and failure injection approach before the implementation specifics are detailed in Chapter 5.

4.1 Research Design

The study follows a controlled experimental comparison methodology [37]. A single, reproducible virtualised testbed hosts both candidate architectures sequentially. The topology, traffic load, WAN bearer characteristics, and failure events are held constant across both phases so that any difference in measured outcomes can be attributed to the architectural change rather than to environmental variation, with one documented exception (probe resolution for the failure-recovery measurement, Section 4.4) discussed explicitly rather than silently treated as identical. This controlled design directly addresses the reproducibility weakness identified in prior simulation-based comparisons [4], where incompatible topologies and traffic models made cross-study conclusions unreliable. To support reproducibility, the complete Containerlab topology definitions, node configuration files, traffic generation scripts, and measurement pipelines developed in this thesis are documented in sufficient detail in Chapter 5 for independent replication on any host capable of running Nokia SR-SIM containers.

The research follows a two-phase experimental design. Phase 1 constructs and evaluates a stitched IS-IS + LDP carrier WAN baseline with EVPN-VXLAN tenant isolation, representative of current carrier-grade DCI practice. This phase establishes ground-truth measurements of steady-state latency, throughput, and failure-recovery time against which Phase 2 is compared. Phase 2 replaces the WAN-transport portion of the stitched architecture with a unified SRv6 uSID fabric, while holding the data centre EVPN-VXLAN fabric identical to Phase 1, and adds TI-LFA fast reroute, Flex-Algo delay-metric path steering, and an OpenDaylight PCE deployed to provide centralised topology visibility; the extent to which that visibility was realised in practice is reported in Chapter 5 and discussed in Chapter 7. Phase 2 is evaluated under identical traffic load, WAN propagation delays, and failure scenarios as Phase 1. A direct cross-architecture measurement campaign under simultaneous load is scoped for future work; the per-phase summaries in Chapters 6 and 7 provide the primary comparative evidence in this thesis.

The phased approach is followed for two reasons. First, deploying and validating each architecture independently before comparing them avoids confounding factors from simultaneous operation. Second, it mirrors the migration path an operator would follow in practice: validate the baseline, then evaluate the replacement under the same conditions.

The independent variable is the WAN-transport architecture: Phase 1 (stitched IS-IS + LDP) versus Phase 2 (unified SRv6 uSID with OpenDaylight PCE), with the data centre fabric held constant in both. The dependent variables are application-layer goodput [38], round-trip latency [39], and traffic recovery time following a link failure event. All other factors, physical link capacity, emulated WAN propagation delays, traffic volume, and MTU, are kept identical across both phases.

A virtualised testbed was chosen over a production deployment for three reasons. First, Nokia SR-SIM [34] runs the same SR OS software image as physical Nokia 7750 SR hardware, so the control-plane behaviour of IS-IS, BGP, EVPN, SRv6, PCEP, and Flex-Algo is faithful to the real platform; data-plane forwarding, however, is software-based, so all timing and throughput figures are relative comparisons between the two architectures rather than measurements representative of physical hardware. Second, the

declarative Containerlab [15] topology files are version-controlled, so any experiment can be reproduced exactly across runs. Third, failure injection can be applied deterministically and reversibly in software without risk to live services.

The WAN segment is deliberately designed with architecturally heterogeneous bearers to reflect real-world DCI deployments. The fiber bearer uses four Nokia SR OS P-routers in a diamond topology, making it an SRv6-aware forwarding mesh in Phase 2 and a plain IP forwarding path in Phase 1. The 5G and satellite bearers are single-hop Linux bridge nodes with `tc netem` delay and jitter configured to match realistic propagation characteristics for each technology class (specific values given in Chapter 5). Because these bearers are Linux containers rather than SR OS nodes, they remain dumb IPv6-only forwarders in both phases and are not SRv6-aware: this reflects the real-world constraint that 5G and satellite access circuits are typically CPE-terminated transport links with no SR capability.

4.2 Evaluation Metrics

Three measurement dimensions capture the DCI performance properties of interest: steady-state latency, which determines whether latency-sensitive workloads can meet their SLAs; throughput, which determines whether the path can sustain bulk data transfer; and failure-recovery time, which determines the operational impact of a WAN link loss on traffic continuity.

Round-trip latency is measured using a dedicated ICMP echo stream of up to 500 probes per measurement, separate from the throughput traffic, so that latency is characterised independently of the bulk flows. Because the testbed is isolated with no competing external load and the conditions are reproducible, within-measurement variance is captured directly in the reported distribution, and the timing-critical failure-recovery result is additionally measured across multiple trials (Section 4.4). Where a result reported in Chapter 6 would benefit from independent replication, this is noted there. Mean, P99, and maximum are reported where the underlying data supports them; standard deviation is reported only where the measurement schema actually captured it, since not every metric in this thesis was logged with per-sample variance. P99 matters more than the mean for DCI applications because tail latency determines the worst-case behaviour of latency-sensitive control-plane traffic crossing the DCI link.

WAN bearer characterisation uses the Nokia SR OS `ping router-instance "Base"` command issued via netmiko, measuring RTT to each bearer's far-end interface. This gives a control-plane view of per-bearer latency independent of host-level traffic flows and application overhead, and serves as ground truth for confirming that emulated WAN delays match the configured `tc netem` values.

DCN bimodal throughput is measured using a custom traffic generator built upon TrafPy [35], which generates a reproducible mix of elephant TCP and mouse UDP flows from a bimodal flow-size distribution with exponential inter-arrival times. 27 elephant TCP flows and 23 mouse UDP flows ($n = 50$ total) run for each measurement window, with a fixed random seed (Section 4.3) so the same flow schedule is used in both phases. Per-class goodput is reported separately, since the two classes have fundamentally different performance characteristics under WAN delay and loss.

Traffic recovery time is the elapsed time between injecting a WAN failure and traffic returning to the normal forwarding path. It is measured at two levels: ICMP probe RTT for the data-plane view, at a probe interval that differs between phases (200 ms in Phase 1, 1 ms in Phase 2, Section 4.4) and is reported as such rather than normalised; and IS-IS adjacency state sampled via netmiko at 2-second intervals for the control-plane view. In Phase 1, recovery requires IS-IS re-convergence followed by LDP re-signalling before the VPRN forwarding path is restored.

4.3 Traffic Generation Strategy

DCI traffic is not uniform. In practice, a DCI link carries a mixture of large, sustained elephant flows such as storage replication, VM migration, and backup jobs alongside small, bursty mouse flows such as health checks, control-plane signalling, and API calls. Measuring goodput with a single constant-bit-rate stream misses this entirely: it says nothing about how the network handles the two classes simultaneously, and the results are not comparable to any published DC traffic study. All traffic experiments therefore use TrafPy with `gen_multimodal_val_dist` and `gen_exponential_dist` to generate a DCN bimodal flow schedule. Flows below 100 KB are treated as mouse flows implemented as UDP; flows above 100 KB are treated as elephant flows implemented as TCP. This produces 27 elephant TCP flows and 23 mouse UDP flows under the configured distribution parameters. Seed 42 is fixed across all runs so that the exact flow

schedule, sizes, and inter-arrival times are identical in every experiment, ensuring that any throughput difference between Phase 1 and Phase 2 is attributable to the architecture and not to traffic variation.

4.4 Failure Injection and Recovery Measurement

Failure injection uses the Nokia SR OS management plane via a netmiko SSH session rather than Linux `tc netem`. SR OS containers use a proprietary I/O stack that bypasses the Linux kernel network layer entirely, making netem loss injection ineffective on SR OS interfaces. Setting `admin-state disable` on the IS-IS-facing interface via the SR OS MD-CLI and issuing a `commit` causes an immediate IS-IS adjacency loss identical in effect to a physical interface-down event. The same command with `admin-state enable` restores the adjacency for recovery measurement.

The evaluation plan for each phase proceeds in the following steps. First, the topology is deployed with `clab deploy` and all control-plane adjacencies are verified before any measurement begins. Second, the steady-state measurement campaign runs the full TrafPy flow schedule with ICMP probing active throughout, aggregating up to 500 probe samples per metric. Third, failure injection is triggered mid-flow by disabling the fiber WAN IS-IS interface on DC1-border, and the recovery time is recorded from both the data-plane ICMP stream and the control-plane adjacency telemetry; for Phase 2's TI-LFA measurement, this step is repeated across 5 failure-injection trials (Chapter 6), since the rapid timescale of fast reroute benefits from repeated sampling. Fourth, the link is restored, and a post-recovery steady-state measurement confirms that the network has returned to baseline. All raw measurement outputs are stored for offline analysis and figure generation.

Chapter 5

Implementation

This chapter describes the concrete implementation of the testbed, the configuration of each experimental phase, and the measurement procedures used to collect the data presented in Chapter 6. Where Chapter 4 establishes what is being compared and why, this chapter covers how it is realised in the Nokia virtualised environment. While specific CLI syntax and feature names reflect Nokia SR OS 25.7.R2, the architectural principles, measurement methodology, and conclusions drawn are not vendor-specific and apply equally to any standards-conformant SR implementation.

The choice of Nokia SR OS 25.7.R2 reflects the industry-supervision context of this research and the availability of SR-SIM as a containerised image running the same software as the physical platform; the same design could in principle be reproduced on any platform supporting the RFC-defined protocol suite. The complete Containerlab topology definitions, node configuration templates, traffic generation scripts, and measurement pipelines are documented in this chapter in sufficient detail for independent replication.

The two phases are deployed and measured sequentially on the same topology rather than simultaneously. This avoids resource contention (both phases compete for the same finite SR-SIM forwarding capacity) and mirrors the migration path an operator would follow in practice: validate the existing stitched baseline under known conditions, then evaluate a candidate replacement under the same conditions. The cross-phase comparison in Section 5.4 is accordingly framed as an analysis of two independently collected datasets.

5.1 Testbed Infrastructure

The testbed runs on a dedicated lab server (Ubuntu 24.04, 32 vCPU, 64 GB RAM). Containerlab v0.75.0 [15] instantiates all nodes as Docker containers, creates virtual Ethernet pairs from a declarative YAML topology file, injects per-node startup configurations, and manages the container lifecycle. The topology file is version-controlled, ensuring exact reproduction of any run.

The topology comprises two simulated data centres (DC1 and DC2), each with a two-tier leaf-spine fabric, connected via three emulated WAN bearers. Each data centre hosts two leaf nodes, a spine, a border gateway, and two traffic endpoints, one attached to each leaf. Phase 1 uses 18 nodes in total; Phase 2 adds the OpenDaylight PCE, bringing the count to 19. The same DC fabric and WAN bearer nodes are used in both phases; only the fiber WAN forwarding plane and border node configurations change between phases. The EVPN-VXLAN control-plane design, leaf-spine topology, and forwarding behaviour are identical across both phases at the configuration level, not only at the topology-diagram level: the EVI, VNI, route-target, and bridging configuration on every leaf and spine are unchanged between the Phase 1 and Phase 2 deployment of either data centre. The Phase 2 fabric is addressed for IPv4 only, consistent with the IPv4 tenant service carried over the unified path, whereas the Phase 1 fabric is dual-stacked; this addressing difference is the only respect in which the leaf configurations differ, and it does not alter the EVPN-VXLAN control plane being held constant. Table 5.1 lists the full node inventory.

Node(s)	Role	Platform
DC1/DC2-leaf, DC1/DC2-leaf2	DC leaf, BGP mac-vrf + IRB	SR Linux IXR-D2L 24.10
DC1/DC2-spine	DC spine, BGP route reflector	SR Linux IXR-D2L 24.10
DC1/DC2-border	DC-WAN border gateway	SR OS sr-1s 25.7.R2
fib-r1..fib-r4	Fiber WAN P-routers (diamond mesh)	SR OS sr-1s 25.7.R2
DC1/DC2-host1, host2	Traffic endpoints (one per leaf)	Alpine Linux 3.19
wan-5g	5G WAN emulator	Alpine + tc netem
wan-satellite	Satellite WAN emulator	Alpine + tc netem
ODL-PCE	OpenDaylight Path Computation Element	ODL 0.18.2

Table 5.1: Testbed node inventory. The DC fabric and WAN bearer nodes are identical across both phases; only the WAN-transport and border configurations differ. The ODL-PCE node is present in Phase 2 only.

A second leaf (`leaf2`) is present in each data centre specifically because a single-leaf topology cannot demonstrate genuine EVPN-VXLAN bridging: with only one leaf acting as the sole VTEP, there is no second bridging endpoint for an EVPN Type-2 route to actually traverse, and any apparent connectivity would not be evidence of EVPN-VXLAN functioning correctly. With `leaf2` present, cross-leaf EVPN-VXLAN bridging within each DC is confirmed via wire capture (Chapter 6), which this thesis treats as the standard of proof for this claim rather than configuration presence alone.

WAN propagation delay is emulated with tc netem [40]: $20\,\text{ms} \pm 5\,\text{ms}$ one-way on the 5G bearer, and $600\,\text{ms} \pm 100\,\text{ms}$ one-way on the satellite bearer, approximating real-world propagation budgets for each technology class. For context, the ITU-R and 3GPP Release 16 frameworks set a stricter 4 ms one-way radio-interface target for enhanced Mobile Broadband (eMBB) [41, 42], while geostationary (GEO) satellite links have a well-established round-trip time of approximately 500–600 ms due to orbital distance [43]. The 600 ms one-way figure configured here yields a round-trip time of roughly 1200 ms, which is deliberately conservative: it bounds the GEO range from above rather than matching it directly, providing a worst-case high-latency bearer against which the transport architectures are compared. As noted in Section 3.6, netem has no effect on SR OS containers, so it is only applied to the Alpine-based 5G and satellite bearer nodes.

5.2 Phase 1: Stitched IS-IS and LDP Baseline

Figure 5.1 shows the complete Phase 1 testbed topology, including the second leaf per data centre discussed above. Two SR Linux data centre fabrics are connected across a multi-bearer WAN domain consisting of the Nokia SR OS fiber P-router mesh and two alternative WAN bearers. The design is symmetric: identical node types and software versions are used in each DC, and the same border node type terminates both the DC fabric and the WAN.

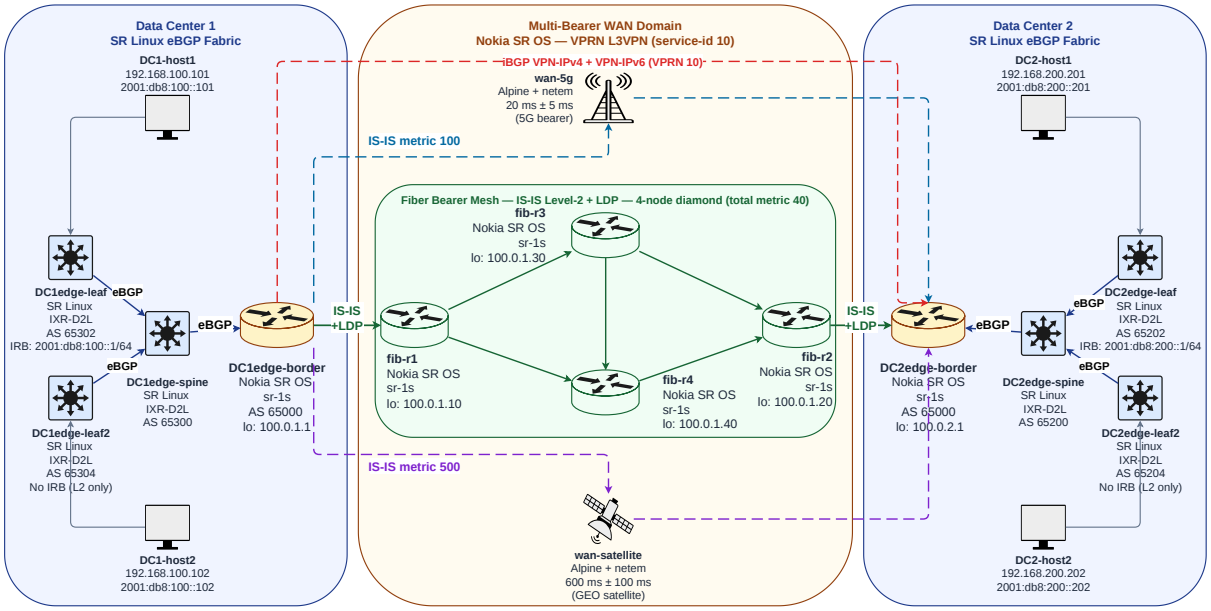


Figure 5.1: Phase 1 testbed topology, including the second per-DC leaf (`leaf2`) added for genuine EVPN-VXLAN bridging. Two SR Linux DC fabrics are connected via a Nokia SR OS fiber WAN diamond (IS-IS Level-2 + LDP, total metric 40) and two alternative WAN bearers (5G, metric 100; satellite, metric 500). The iBGP VPN-IPv4 session between border nodes carries VPRN 10 L3VPN routes over LDP-signalled paths.

DC Fabric. Both DCs run a two-tier leaf-spine fabric on Nokia SR Linux, with two leaves per DC (`leaf`, `leaf2`) acting as independent VTEPs. One traffic host attaches to each leaf, so cross-DC traffic between the sending host in DC1 and the receiving host in DC2 traverses a different leaf at each end and therefore genuinely crosses the EVPN-VXLAN fabric rather than being locally bridged. Each leaf node runs a mac-vrf network instance with an IRB interface that provides the default gateway for attached hosts; an `advertise-neighbor-type router-host` setting is applied on `leaf2` specifically, since it is incompatible with the IRB configuration present on `leaf1`. The spine nodes act as BGP route reflectors for the DC underlay, with each leaf and spine in a separate autonomous system following the per-leaf eBGP underlay design recommended for SR Linux DC fabrics [44].

WAN Transport. The fiber WAN nodes run IS-IS Level-2 with LDP [45] for label distribution, arranged in a diamond topology. There are no SR extensions, no traffic engineering extensions, and no fast reroute in Phase 1. LDP distributes labels hop-by-hop based on IS-IS shortest paths. IS-IS metrics are configured asymmetrically so that the fiber diamond is always preferred over 5G and satellite, preventing unintended ECMP across bearers.

Service Layer. A VPRN service on each border node [26] carries DC-to-DC tenant traffic, resolving the iBGP VPN-IPv4 next-hop via LDP label-switched paths. An iBGP session between border loopbacks carries VPN-IPv4 routes, and eBGP within the VPRN connects each border to its upstream DC spine. The dual-stack extension to this service (6VPE, Chapter 3) carries cross-DC host-to-host traffic over both address families on the Phase 1 stitched transport: the IPv6 primary path and the IPv4 backup path are both confirmed working end-to-end (Chapter 6, Section 6.1). The Phase 2 unified SRv6 path is configured for the IPv4 tenant service (End.uDT4) only; bringing the equivalent cross-DC IPv6 service up over the unified path was out of scope for this iteration and is identified as future work (Section 7.5).

Failure Injection. Link failures are injected by disabling the IS-IS interface on DC1-border toward the fiber WAN via the SR OS management plane using netmiko. This causes an immediate IS-IS adjacency loss, correctly simulating a physical fiber access failure. The management-plane approach is necessary because Nokia SR OS containers use a userspace I/O stack that bypasses Linux `tc netem` entirely.

5.2.1 Phase 1 Evaluation Procedure

The Phase 1 evaluation proceeds in four steps. First, the topology is deployed, and a 120-second stabilisation period allows IS-IS to converge across all SR OS nodes; IS-IS adjacencies on DC1-border and end-to-end host reachability are verified before any measurement begins. Second, the steady-state mea-

surement runs the full TrafPy bimodal flow schedule with ICMP probing active throughout, collecting RTT histograms (up to 500 samples) and per-class goodput. Third, bearer characterisation probes are issued from each border node to the far-end interface of each WAN bearer, confirming that emulated delays match the configured netem values. Fourth, failure injection is triggered mid-flow by disabling the fiber IS-IS interface; ICMP probes at 200 ms intervals capture the data-plane recovery timeline while IS-IS adjacency state is polled at 2-second intervals for the control-plane view. The link is subsequently restored and a post-recovery steady-state window confirms return to baseline.

5.3 Phase 2: Unified SRv6 uSID WAN Transport

The second phase replaces the WAN-transport portion of the stitched architecture, the fiber P-router mesh and the WAN-facing function of the border nodes, with a unified SRv6 uSID data plane. The data centre fabric described above (`leaf`, `leaf2`, spine, EVPN-VXLAN, mac-vrf, IRB) is unchanged from Phase 1; this is the controlled variable underpinning this thesis's comparison (Chapter 4), and Figure 5.2 marks the DC-fabric portion of the diagram explicitly as unchanged for this reason. The border nodes are reconfigured to terminate VPRN tenant traffic into an SRv6 uSID-tunnelled path across the WAN, rather than into an LDP label-switched path; this is the only change made at the border, and EVPN-VXLAN tenant isolation within the DC fabric is not touched.

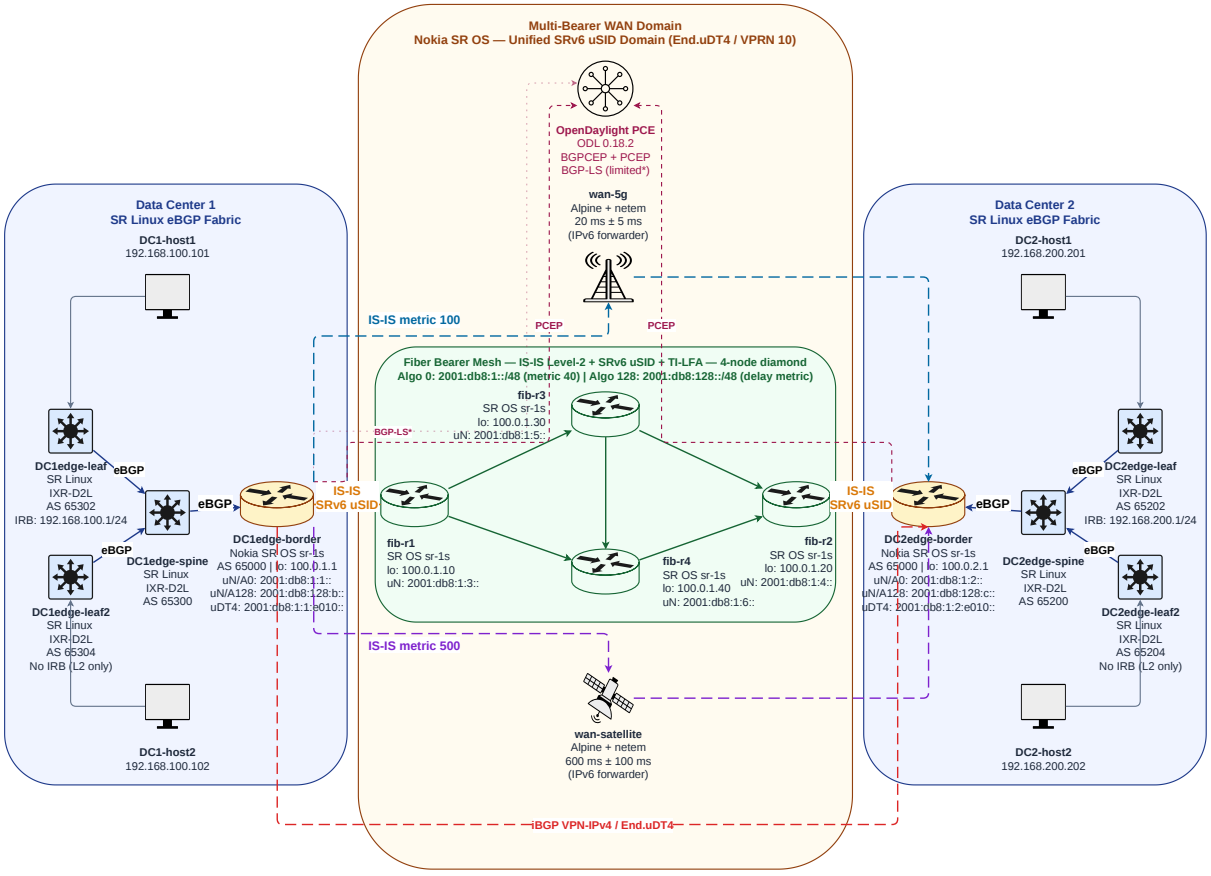


Figure 5.2: Phase 2 testbed topology. The data centre fabric (`leaf`, `leaf2`, spine, EVPN-VXLAN) is unchanged from Phase 1 and is marked as such. The fiber WAN diamond is replaced by a unified SRv6 uSID domain with TI-LFA and dual Flex-Algo locators. The OpenDaylight PCE connects to both border routers via PCEP (management network, port 4189) and BGP-LS (dedicated data-plane interfaces, port 179). PCEP sessions are fully synchronised with stateful and PCE-initiated LSP capabilities; BGP-LS sessions are established with LINKSTATE address family negotiated (Section 5.3). The iBGP label between borders reflects VPN-IPv4/End.uDT4 only, since cross-DC host-to-host IPv6 over the unified path was not verified working end-to-end (Section 7.5).

SRv6 uSID Configuration. Each border node is assigned a /48 uSID locator block. The End.uN

micro-SID provides node-level forwarding, End.uA provides adjacency-level steering for Flex-Algo path enforcement, and End.uDT4 provides per-VRF IPv4 table lookup for L3VPN service delivery, replacing the LDP-based label-switched path of Phase 1 for the WAN leg specifically; VPRN tenant isolation logic itself is unchanged. All fiber P-routers participate in the IS-IS domain and forward SRv6 traffic based on the outer IPv6 destination address without inspecting the SRH. The 5G and satellite bearers remain dumb IPv6-only forwarders as in Phase 1.

IS-IS Domain. A single IS-IS Level-2 domain spans the WAN-transport SR OS nodes. IS-IS TE metric extensions [14] advertise unidirectional link delay on all fiber interfaces, providing the delay sub-TLVs consumed by Flex-Algo 128. BGP-LS [33] is configured to export this IS-IS topology to the OpenDaylight PCE via dedicated data-plane interfaces (10.99.1.0/30 and 10.99.2.0/30) using eBGP (AS 65000 on SR OS, AS 65001 on ODL) with the LINKSTATE address family. The SR OS border routers are configured with `link-state-route-export true`, `database-export` with `bgp-ls-identifier` and `igp-identifier`, and an accept-all export policy on the ODL peer group.

TI-LFA. TI-LFA fast reroute is enabled on all WAN-transport IS-IS interfaces. The repair path bypasses the primary fiber link via the alternate P-router path, pre-computed during IS-IS initialisation and pre-installed in the forwarding plane. No control-plane processing is required at failure time. How fast activation actually is on this testbed is reported as a measured result in Chapter 6 (mean 1.2 ms, maximum 3 ms across 5 failure-injection trials) rather than asserted here in advance.

Flex-Algo and PCE. Flex-Algo 0 operates as the default IGP topology using standard IS-IS metrics. Flex-Algo 128 is configured with delay metric type, computing minimum-latency paths across the fiber diamond using the advertised TE delay sub-TLVs. Both border nodes carry dual uSID locators, one per algorithm.

The OpenDaylight PCE is deployed as a containerised instance (`opendaylight:0.18.2`) connected directly to both border routers. PCEP uses the management network: each SR OS border peers with ODL at 172.20.20.50 on port 4189, with `route-preference outband` to direct the TCP session through the management routing instance. Both PCEP sessions achieve full synchronisation with stateful and PCE-initiated LSP capabilities negotiated.

BGP-LS uses dedicated data-plane interfaces: ODL (AS 65001) actively connects to each SR OS border (AS 65000, configured as `passive true`) on port 179. Both BGP-LS sessions reach Established state with the LINKSTATE address family negotiated. The BGP peer configuration is pushed via the ODL RESTCONF API, with peer-type, AS number, transport, and AFI/SAFI specified directly on each neighbor.

The PCE evaluation is therefore scoped to establishing and validating this control plane, the layer on which any subsequent path computation depends; end-to-end SRv6 path instantiation is the next stage and is treated as such. Flex-Algo path differentiation is confirmed to operate independently of the PCE.

5.3.1 Phase 2 Evaluation Procedure

The Phase 2 evaluation follows the same four-step structure as Phase 1. After deployment, IS-IS adjacency across all nodes, SRv6 uSID SID advertisement via `show router segment-routing-v6 micro-segment-local-sid`, and PCEP session establishment are verified before measurement begins. The same TrafPy bimodal flow schedule with seed 42 is then run under steady-state conditions, matching Phase 1's measurement structure. Failure injection uses the same netmiko management-plane approach as Phase 1, disabling the same IS-IS interface on DC1-border; this step is repeated across 5 failure-injection trials specifically for the TI-LFA recovery measurement, since the rapid timescale of fast reroute benefits from repeated sampling. Post-recovery steady-state measurement completes the campaign.

5.4 Cross-Phase Comparison

Both architectures are subjected to identical failure and load scenarios by deploying the Phase 1 and Phase 2 configurations sequentially on the same Containerlab topology, rather than as a third independently-deployed experimental phase. The comparison figures generated from this pairing, fiber RTT distribution, bearer RTT comparison, bimodal throughput, failure-recovery timeline, and IS-IS convergence telemetry, are presented together in Chapter 6 specifically so that any observed difference can be read against both phases' data at once, but they are an analysis of the two phases' independently-collected results rather than a separately-run measurement campaign of their own.

Chapter 6

Results

This chapter presents the measurement results for both experimental phases. Phase 1 establishes the IS-IS + LDP stitched baseline and Phase 2 presents the unified SRv6 uSID fabric results, with a direct cross-phase comparison at the end of each relevant section. All figures and tables in this chapter reflect the most recent measurement run for each phase; values reported here supersede any earlier draft figures.

6.1 Phase 1: IS-IS and LDP Stitched Baseline

6.1.1 Fiber Bearer RTT

Figure 6.1 shows the RTT distribution between DC1-host and DC2-host across the fiber diamond WAN, on the working IPv6 primary path. The mean RTT is 3.70 ms with a P99 of 5.45 ms ($n = 471$).

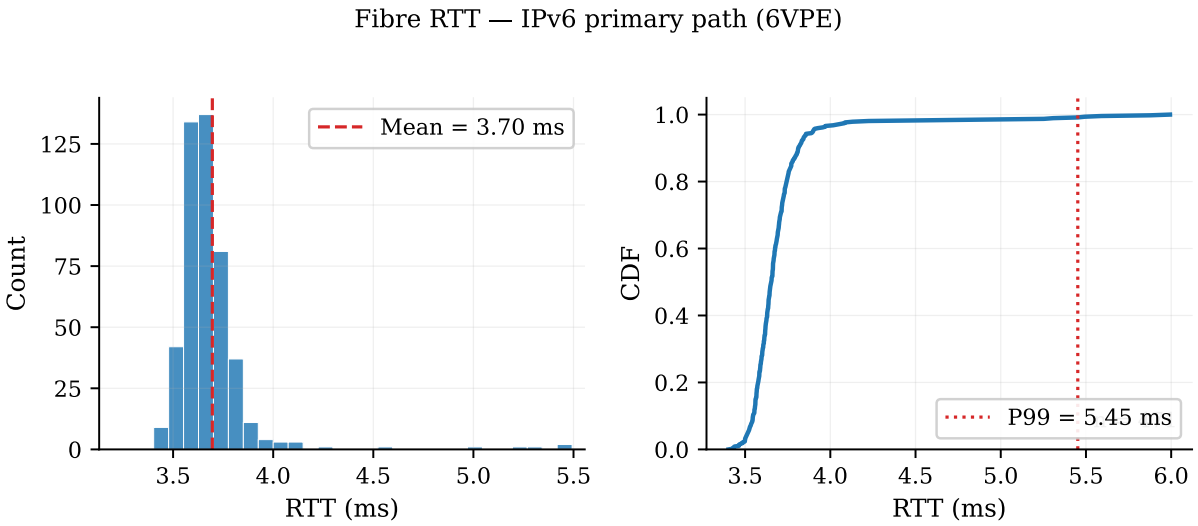


Figure 6.1: Phase 1 fiber bearer RTT, IPv6 primary path ($n = 471$). Left: histogram. Right: CDF with P99 = 5.45 ms.

The mean of 3.70 ms is consistent with SR OS software-forwarding latency accumulated across the four-hop fiber diamond, since netem is not applied to the SR OS fiber path (Section 3.6). A separate measurement on the IPv4 backup path gives a comparable mean of 3.65 ms (P99 4.05 ms, $n = 500$); the small difference between the two address families is attributable to the additional VPN-IPv6 processing on the working IPv6 path rather than to a difference in physical path length, since both paths traverse the same physical route. Throughout this measurement, the IS-IS route table on DC1-border consistently showed the fiber diamond path as active, confirming all probes traversed the intended path.

6.1.2 WAN Bearer Characterisation

Figure 6.2 shows the RTT for each of the three WAN bearers measured from DC1-border via the SR OS management plane.

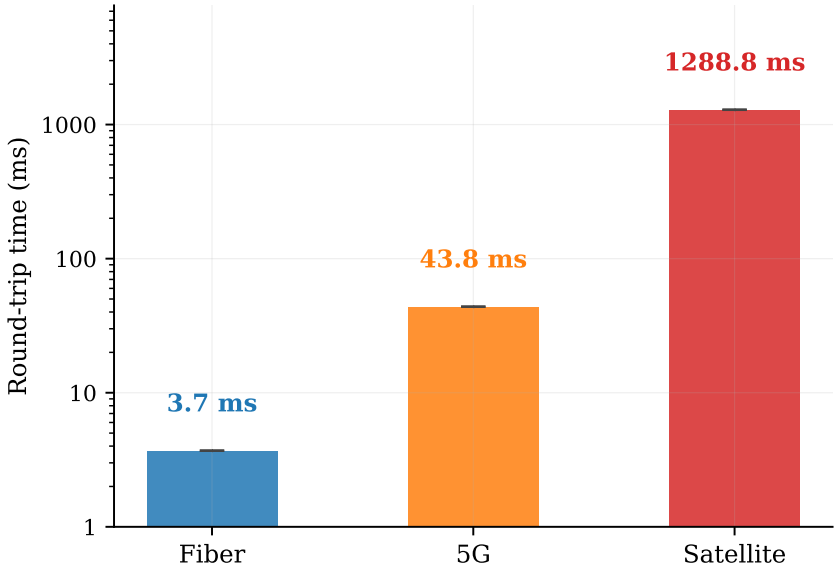


Figure 6.2: Phase 1 WAN bearer RTT. Log scale. Fiber: 3.70 ms. 5G: 43.8 ms (min 34.8, max 86.4, $n = 20$). Satellite: 1288.8 ms (min 1070, max 2348, $n = 10$). Error bars show min–max range.

The three-decade log-scale span illustrates the fundamental heterogeneity of the three bearers. Switching from fiber to 5G multiplies RTT by approximately 12×; satellite is approximately 350× worse. The satellite figure is consistent with the configured 600 ms ± 100 ms one-way netem delay once doubled into a round-trip figure, plus queuing overhead. In Phase 1, bearer selection is governed entirely by manually configured IS-IS metrics with no dynamic path selection based on measured delay.

6.1.3 DCN Bimodal Traffic

Figure 6.3 shows the TrafPy DCN bimodal results with seed 42 and $n = 50$ flows (27 elephant TCP, 23 mouse UDP).

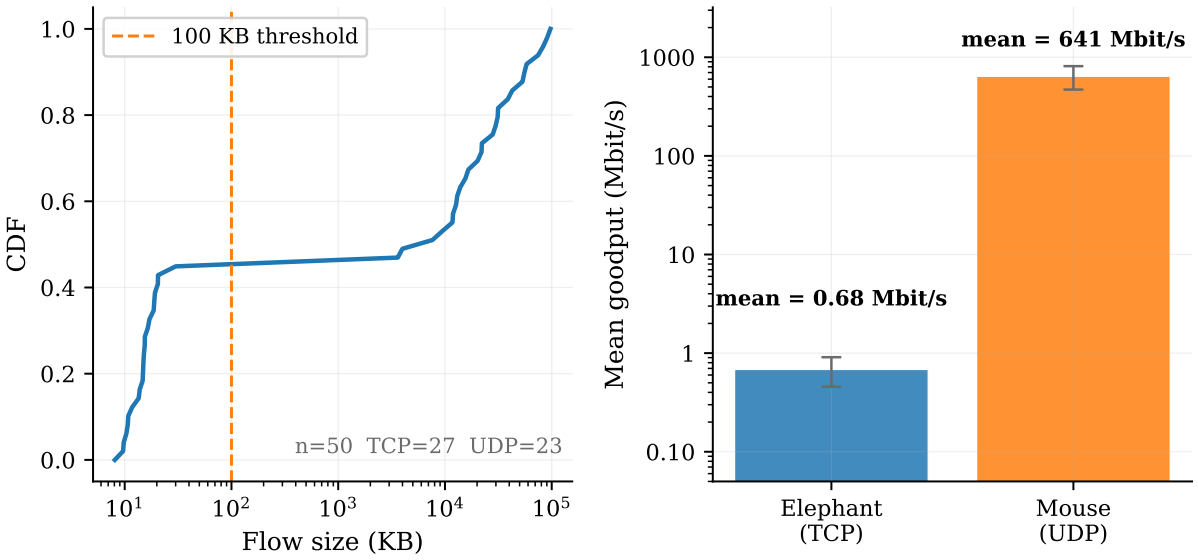


Figure 6.3: Phase 1 TrafPy DCN bimodal traffic ($n = 50$ flows, seed 42). Left: flow-size CDF; the dashed vertical line marks the 100 KB threshold separating mouse (UDP) from elephant (TCP) flows. Right: per-class goodput; error bars indicate one standard deviation across flows within each class.

Mouse UDP flows achieved a mean goodput of 641 ± 171 Mbit/s with 100% completion across all 23 flows. Elephant TCP flows achieved a mean goodput of 0.68 ± 0.23 Mbit/s. This is an important distinction to state precisely: inspection of the raw output shows that TCP connections *do* establish (the SYN exchange completes across the VPRN path), but throughput then stalls at a very low rate rather than ramping up – this is a different failure mode from connection establishment failure, and is consistent with an MTU interaction between the host interface and the VXLAN-over-LDP encapsulation stack causing repeated retransmission or severe MSS-driven throttling rather than an outright handshake failure. UDP flows are largely unaffected because they do not depend on TCP's congestion/retransmission behaviour. The same low elephant-flow throughput recurs in Phase 2 under an architecturally different WAN transport (Section 6.2.3), which is part of the evidence used in Chapter 7 to argue that this is a shared testbed constraint rather than a Phase 1-specific architectural defect. It is discussed further in Section 7.7.

6.1.4 Failure Recovery

The failure scenario disables the IS-IS interface toward fib-r1 on DC1-border via the SR OS management plane, removing the fiber WAN entry from the IS-IS LSDB and forcing SPF re-computation followed by LDP label re-signalling on the surviving 5G path.

Figure 6.4 shows the ICMP probe time series during the reconvergence measurement, at 200 ms probe resolution. RTT remains at the fiber baseline for approximately 4.07 s after failure injection, then steps to approximately 44 ms as the 5G bearer becomes active. The interface is subsequently restored, and RTT returns to fiber levels within two probe intervals. The 200 ms probe interval used in Phase 1 bounds the convergence-time measurement to ±200 ms resolution; a much finer, per-millisecond characterisation is provided by the Phase 2 TI-LFA measurement in Section 6.2.5, and the two measurements use different probe resolutions for reasons discussed in Chapter 4.

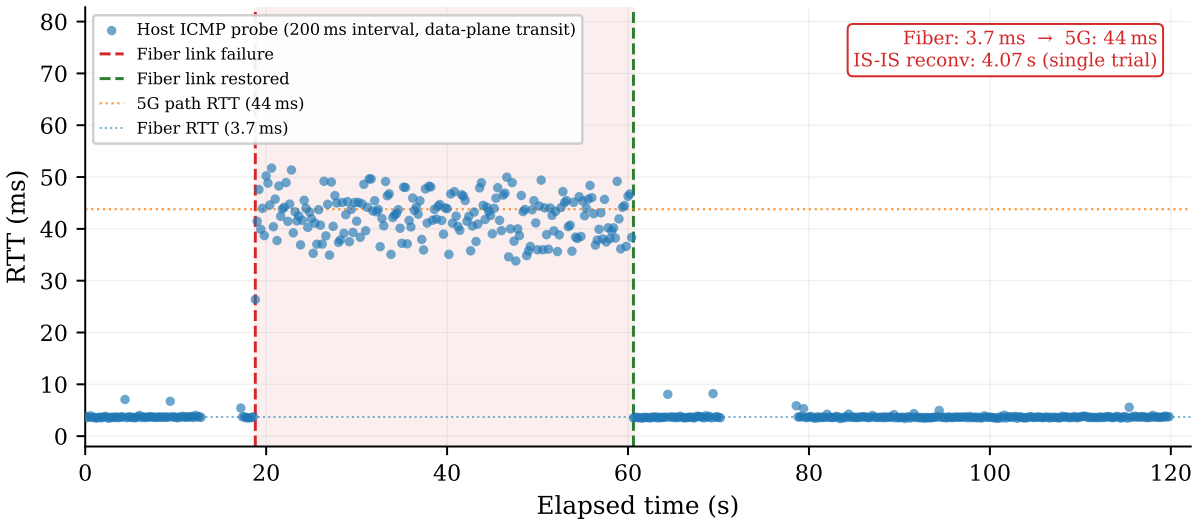


Figure 6.4: Phase 1 ICMP probe RTT during IS-IS interface failure (200 ms probe interval). Red dashed: failure injection. Green dashed: interface restoration. Convergence window: 4.07 s.

The IS-IS control-plane telemetry in Figure 6.5 corroborates the data-plane observation. The figure shows the adjacency count on DC1-border dropping immediately on failure, with the 5G and satellite adjacencies remaining intact and returning to their original value on restoration, consistent with the data-plane RTT step observed in Figure 6.4.

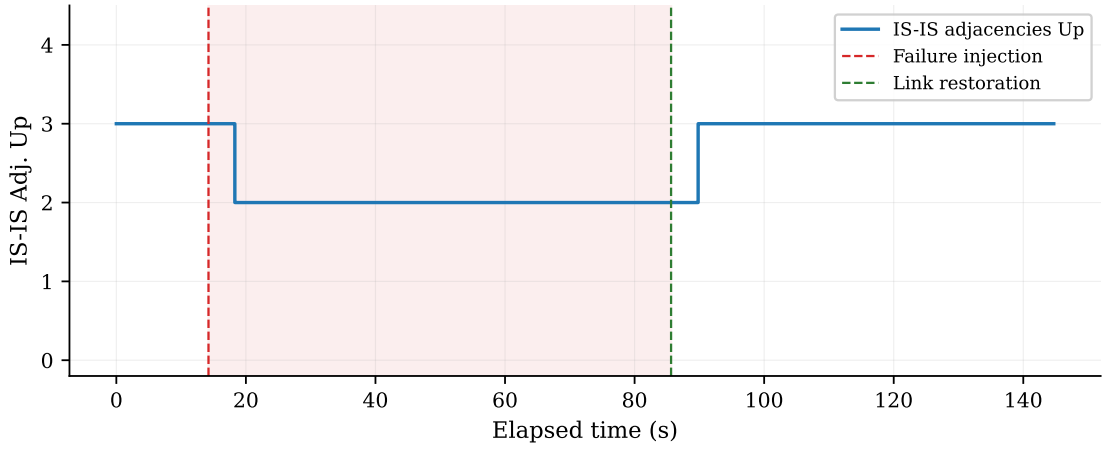
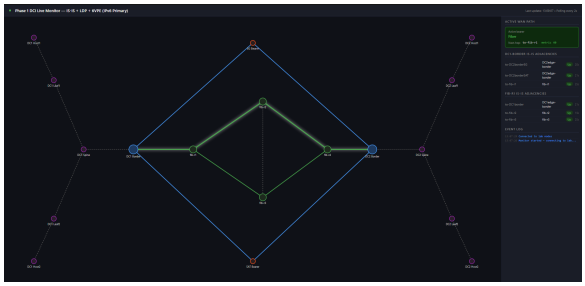
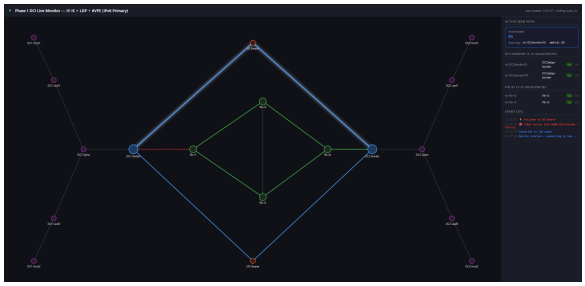


Figure 6.5: Phase 1 IS-IS control-plane telemetry during the failure event. Red dashed: failure. Green dashed: restoration. The adjacency count drops on failure and is restored on link restoration.

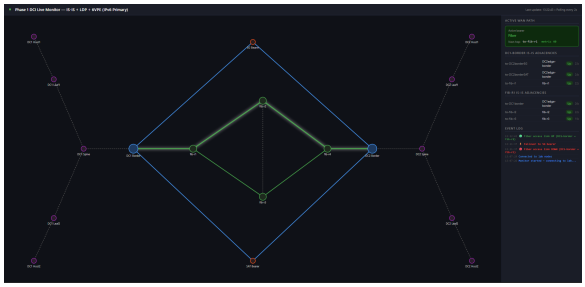
The live monitor screenshots in Figure 6.6 show the topology state at key points in the failure sequence, confirming that the visual adjacency state, active bearer, and event log are consistent with the telemetry data.



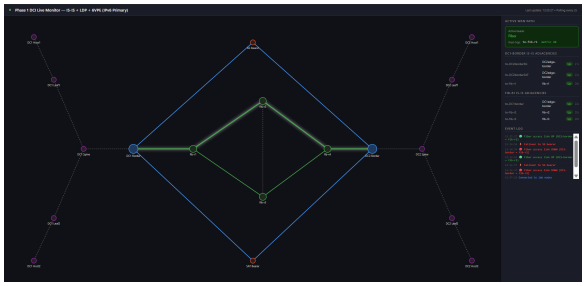
(a) Steady state: fiber active, all IS-IS adjacencies Up on both DC1-border and fib-r1.



(b) Failure injected: DC1-border → fib-r1 link shown as red dashed, active bearer switches to 5G.



(c) Recovery: link returns to green, fiber active, adjacency count restored.



(d) Final state with complete event log showing the failure-and-recovery cycle.

Figure 6.6: Phase 1 live monitor (IS-IS + LDP) during failure injection and recovery. Green links indicate IS-IS adjacency Up; red dashed link indicates adjacency Down.

6.1.5 Phase 1 Summary

Three findings characterise Phase 1. First, steady-state latency is low and consistent across both address families: 3.70 ms mean fiber RTT (IPv6 primary) and 3.65 ms (IPv4 backup) are both in line with the configured WAN propagation delay, and the three-decade span across bearer types confirms the multi-bearer heterogeneity is operating as designed. Second, throughput is strongly class-dependent: mouse UDP flows achieve 641 Mbit/s mean goodput while elephant TCP flows stall at 0.68 Mbit/s despite successfully establishing connections, a pattern consistent with an MTU/MSS interaction discussed in Chapter 7. Third, and most importantly for the cross-phase comparison: the IS-IS + LDP reconvergence window of 4.07 s, measured at 200 ms probe resolution, constitutes the stitched baseline against which Phase 2 TI-LFA is compared in Section 6.3.

6.2 Phase 2: Unified SRv6 uSID Fabric

Phase 2 replaces the WAN-transport portion of the stitched architecture with a unified SRv6 uSID domain, while the data centre fabric is unchanged from Phase 1 (Chapter 5). Before any measurement, the data plane is verified: End.uDT4 is confirmed active, 7 VPRN tenant routes are confirmed reachable via SRv6 tunnelling, 16 SRv6 tunnels are present, and 0 MPLS tunnels remain, confirming that LDP has been fully replaced by SRv6 for the WAN leg. The active uSID state visible in the live monitor confirms uN Algo0 (2001:db8:1:1::), uN Algo128 (2001:db8:128:b::), and the uDT4 SID are all advertised and active on DC1-border.

6.2.1 Fiber Bearer RTT

Figure 6.7 shows the host-to-host RTT distribution across the same fiber diamond path. The mean RTT is 4.13 ms (std 0.681 ms) with a P99 of 6.96 ms ($n = 500$, 0% loss).

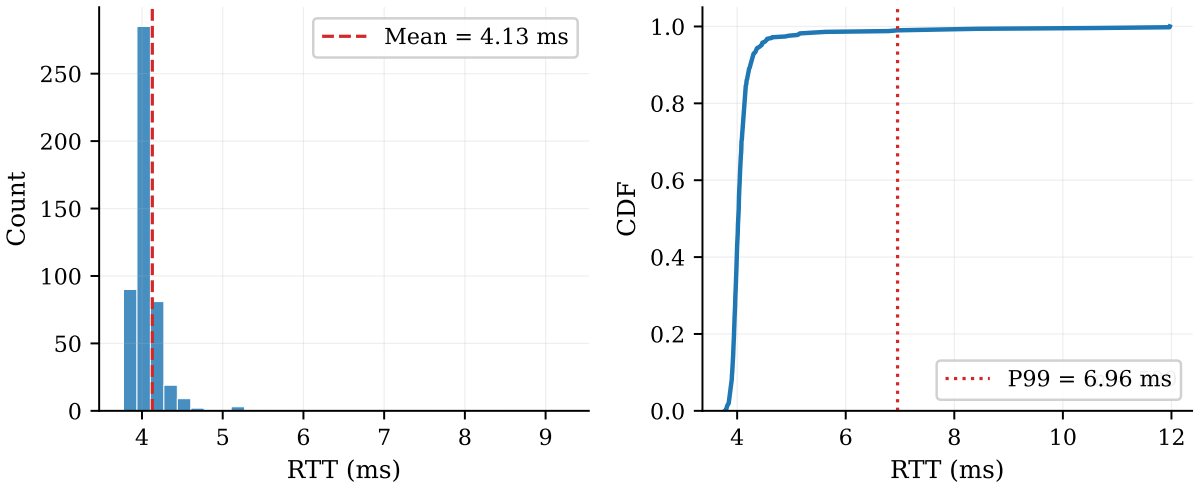


Figure 6.7: Phase 2 fiber bearer RTT, host-to-host ($n = 500$). Left: histogram. Right: CDF with P99 = 6.96 ms.

The mean RTT of 4.13 ms is measurably higher than Phase 1's IPv6-primary-path mean of 3.70 ms. This thesis reports this difference as a measured fact without asserting a specific mechanism for it: while SRv6 uSID encapsulation/decapsulation processing at the border PEs is a plausible contributor, this was not isolated by a dedicated measurement, and attributing the full difference to that single cause would overstate what was actually tested. This is discussed further in Chapter 7.

6.2.2 WAN Bearer Characterisation

A separate, router-originated measurement, methodology-matched to the Phase 1 bearer probes rather than to the host-to-host metric above, gives the per-bearer RTT shown in Figure 6.8: fiber 2.477 ms (std 0.091 ms, $n = 30$), 5G 40.52 ms (std 3.278 ms, $n = 30$), and satellite 1199.4 ms (std 105 ms, $n = 10$). These figures should not be subtracted from the host-to-host figure above to infer a host-to-router latency component, since the two were measured with different methodologies (router-originated ICMP vs. host-pair ICMP).

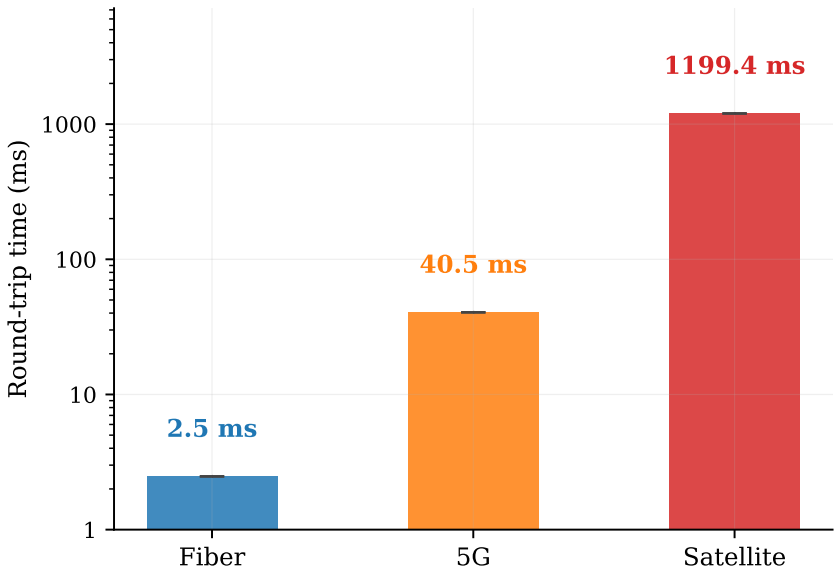


Figure 6.8: Phase 2 WAN bearer RTT, router-originated. Log scale. Fiber: 2.477 ms. 5G: 40.52 ms. Satellite: 1199.4 ms.

The 5G and satellite router-originated values are close to the Phase 1 bearer-characterisation figures, consistent with these bearers being dumb IPv6-only forwarders in both phases with no architectural

change between them.

6.2.3 DCN Bimodal Traffic

Figure 6.9 shows the Phase 2 TrafPy results under the identical seed 42 schedule. Mouse UDP flows achieved a mean goodput of 706 Mbit/s (std 194 Mbit/s) with 100% completion. Elephant TCP flows achieved a mean goodput of 0.68 Mbit/s (std 0.19 Mbit/s).

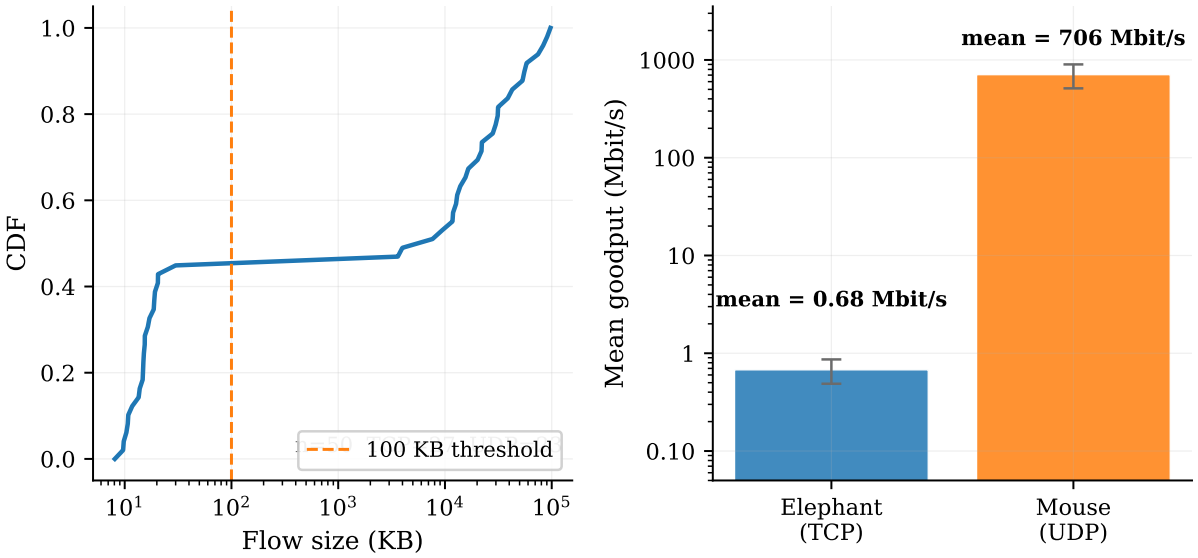


Figure 6.9: Phase 2 TrafPy DCN bimodal traffic ($n = 50$ flows, seed 42). Left: flow-size CDF; the dashed vertical line marks the 100 KB threshold separating mouse (UDP) from elephant (TCP) flows. Right: per-class goodput on log scale; error bars indicate one standard deviation across flows within each class.

Both the mouse and elephant figures are statistically close to their Phase 1 counterparts (641 Mbit/s and 0.68 Mbit/s, respectively), supporting the claim made in Section 6.1 that the elephant-flow throughput ceiling is a shared testbed constraint rather than a property of either WAN-transport architecture: an architecture-specific bottleneck would be expected to produce a phase-dependent difference, and none of that magnitude is observed here.

6.2.4 WAN-Leg Encapsulation Overhead

The per-packet WAN-leg encapsulation overhead of the two architectures was measured directly from packet captures taken on the fiber WAN-facing interface of the DC1 border node, with an identical inner payload in both phases (an ICMP echo carrying a 56-byte data field: a 20-byte IP header, an 8-byte ICMP header, and a 56-byte payload gives an 84-byte inner IPv4 packet). Capturing the same traffic direction at the same point in both phases isolates the encapsulation difference from any payload variation. Table 6.1 reports the resulting on-the-wire frame sizes.

Component	Phase 1 (MPLS)	Phase 2 (SRv6 uSID)
Ethernet header	14 B	14 B
WAN-leg encapsulation	8 B	40 B
Inner IPv4 + ICMP	84 B	84 B
Frame on wire	**106 B**	**138 B**

Table 6.1: Measured WAN-leg per-packet encapsulation overhead, captured on the fiber WAN interface for an identical 84-byte inner packet. Phase 1 carries a two-label MPLS stack (8 B); Phase 2 carries a 40-byte outer IPv6 header with the uSID in the destination address and no Segment Routing Header (confirmed by the IPIP next-header value, i.e. reduced encapsulation). The unified transport is 32 bytes heavier per packet at this single-hop scale.

The Phase 1 MPLS frame measures 106 bytes on the wire, carrying a two-label stack (a transport label and a service label). The Phase 2 SRv6 uSID frame measures 138 bytes, carrying a single 40-byte outer IPv6 header; the packet capture confirms that no Segment Routing Header is present, since the uSID is encoded in the IPv6 destination address itself under reduced encapsulation. With the inner payload held identical, the 32-byte difference is therefore purely the WAN-leg encapsulation cost: at this testbed's single-hop WAN scale, the unified SRv6 uSID transport imposes 32 bytes more per packet than the legacy MPLS label stack. This is the expected direction at a single hop, where a 40-byte IPv6 outer header is unavoidable while MPLS needs only an 8-byte label stack; uSID's header-compression advantage is a multi-hop property (Chapter 3) that this topology does not exercise, and is discussed further in Chapter 7.

6.2.5 TI-LFA Failure Recovery

Figure 6.10 shows the TI-LFA failure-recovery results across 5 failure-injection trials, at 1 ms probe resolution. The data-plane packet-loss window T_{black} has a mean of 1.2 ms (std 1.2 ms), with a minimum of 0 ms and a maximum observed value of 3 ms across the 5 trials. The backup-path RTT, RTT_{bkp}, is stable at 40.1 ± 0.1 ms across all trials, consistent with traffic being carried over the 5G bearer during the repair window. The elevated-RTT duration, T_{elev}, averages 13.1 ± 0.04 s, reflecting the much slower, separate process of full IS-IS reconvergence back onto the restored primary fiber path once it is available again – this is a distinct event from T_{black} and should not be confused with it.

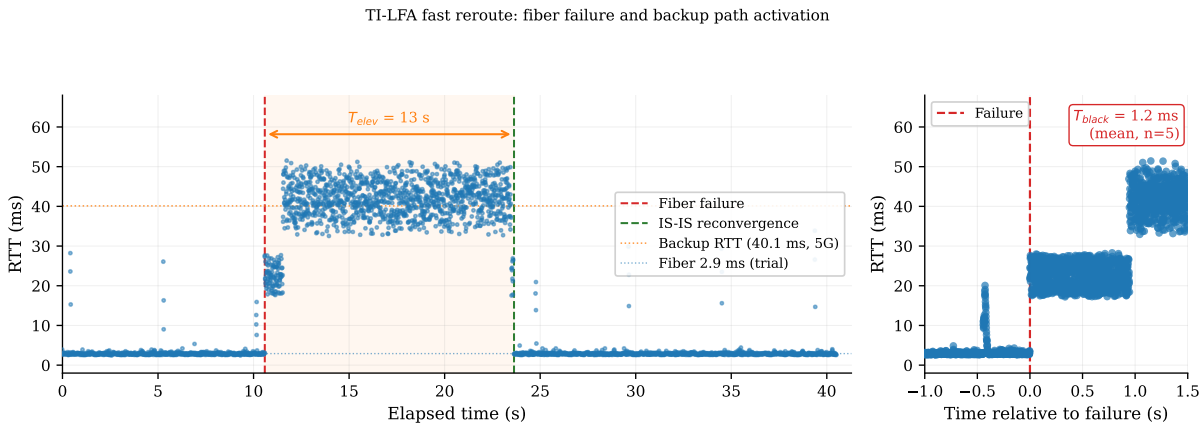
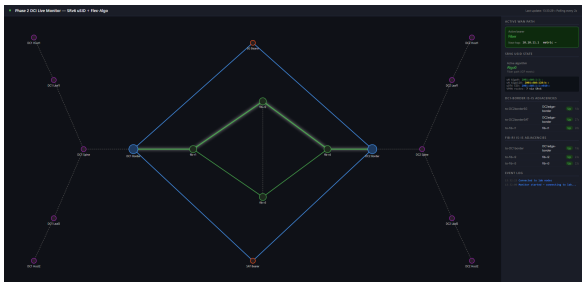


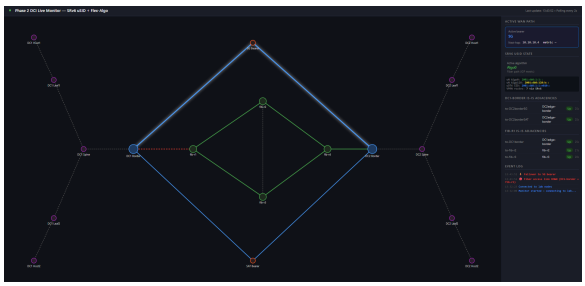
Figure 6.10: Phase 2 TI-LFA failure recovery across 5 failure-injection trials (representative trial shown for the elapsed-time panel). Left: T_{elev} for the representative trial. Right: T_{black} per trial, mean = 1.2 ms, max = 3 ms.

The small variance in T_{black} across trials (0–3 ms) is consistent with the 1 ms probe granularity used for this measurement specifically (distinct from the 200 ms resolution used for the Phase 1 measurement, Chapter 4) and the timing of the failure event relative to the probe schedule. T_{black} at this magnitude is orders of magnitude faster than the 4.07 s window measured in Phase 1; this comparison, and its appropriate qualifications, is discussed in full in Chapter 7.

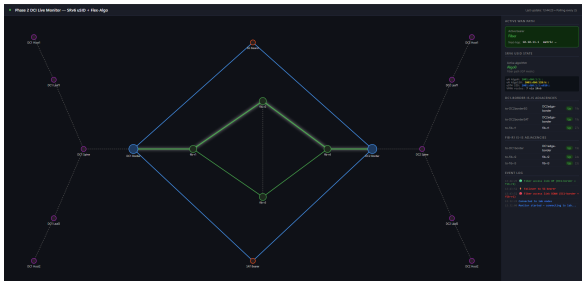
The live monitor screenshots in Figure 6.11 confirm the SRv6 uSID state remains consistent throughout the failure sequence: the active algorithm, the uN and uDT4 SIDs, and the VPRN route count (7 routes via SRv6) are unchanged across steady-state, failure, and recovery, since TI-LFA activates the pre-computed repair path without any SID reconfiguration.



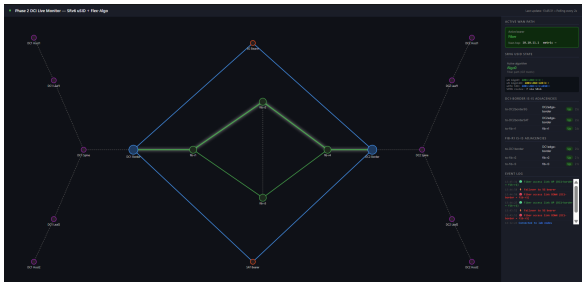
(a) Steady state: fiber active, Algo0, uSID SIDs advertised, 7 VPRN routes via SRv6.



(b) Failure injected: DC1-border → fib-r1 link red dashed, active bearer switches to 5G. SRv6 uSID state unchanged.



(c) Recovery: fiber restored, all adjacencies Up, active bearer returns to Fiber.



(d) Final state with complete event log across trials, confirming deterministic TI-LFA activation.

Figure 6.11: Phase 2 live monitor (SRv6 uSID + Flex-Algo) during TI-LFA failure trials. Green links indicate IS-IS adjacency Up; red dashed link indicates adjacency Down.

6.2.6 Phase 2 Summary

Three findings characterise Phase 2. First, steady-state performance is comparable to Phase 1: the host-to-host fiber RTT of 4.13 ms is approximately 12% higher than Phase 1's 3.70 ms (discussed further in Chapter 7), and throughput is statistically indistinguishable between phases, with the same elephant-flow ceiling present in both. Second, and centrally, TI-LFA bounds the data-plane disruption window to a mean of 1.2 ms (maximum 3 ms across 5 failure-injection trials), orders of magnitude shorter than the Phase 1 reconvergence window; the elevated-RTT period T_{elev} (13.1 s mean) reflects the separate IS-IS reconvergence process and should not be confused with the TI-LFA local-repair time. Third, SRv6 verification confirms full LDP replacement at the WAN leg (7 VPRN routes carried via SRv6, 16 SRv6 tunnels, 0 MPLS tunnels), confirming the unified data plane is operational.

6.3 Cross-Phase Comparison

Each phase was evaluated independently before any cross-phase comparison was attempted. This ordering is deliberate: it mirrors the migration path an operator would follow in practice (validate the existing baseline, then evaluate the replacement under the same conditions) and ensures each architecture is fully understood on its own terms before differences are attributed to the change between them. The cross-phase analysis in this section does not introduce new measurements; it places the per-phase results from Sections 6.1 and 6.2 on shared axes so that each metric can be read comparatively in one place.

6.3.1 Steady-State RTT Comparison

Figure 6.12 overlays the fiber RTT distribution and CDF from both phases. Phase 2 host-to-host fiber RTT is approximately 12% higher than Phase 1's IPv6-primary-path RTT (4.13 ms vs. 3.70 ms). This comparison uses Phase 1's IPv6-primary-path RTT (3.70 ms, P99 5.45 ms), matching the Phase 1 headline figure used throughout this thesis (Section 6.1), so the figure is directly consistent with the rest of the chapter rather than introducing a separate Phase 1 reference point.

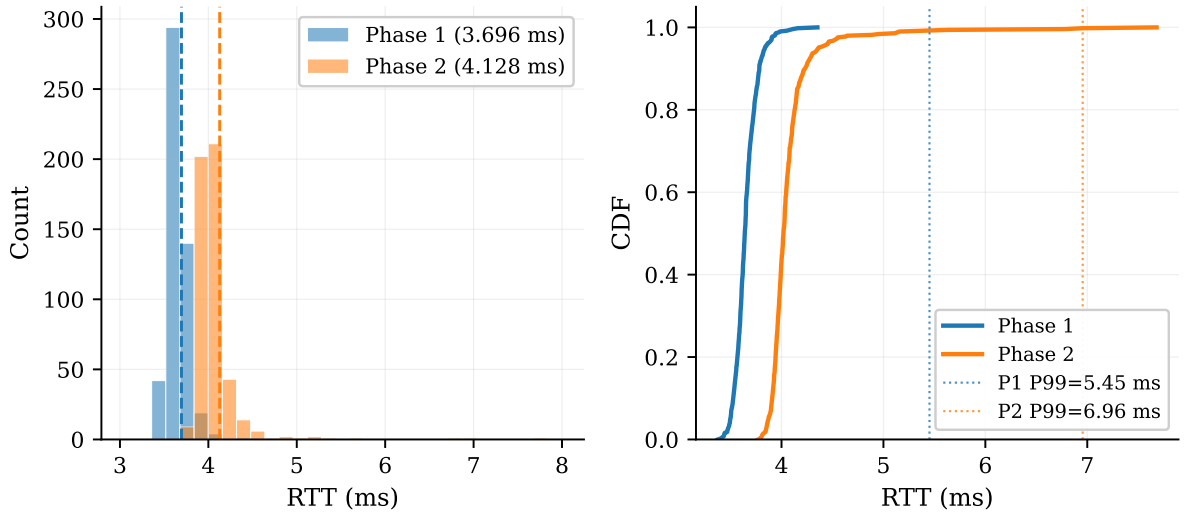


Figure 6.12: Cross-phase fiber RTT comparison. Left: histogram of Phase 1 IPv6-primary-path RTT (3.70 ms) vs. Phase 2 host-to-host RTT (4.13 ms). Right: CDF with P99 lines, Phase 1 5.45 ms vs. Phase 2 6.96 ms.

6.3.2 WAN Bearer RTT Comparison

Figure 6.13 places all three WAN bearers from both phases on a shared log-scale axis. 5G and satellite bearer RTT are marginally lower in Phase 2 (40.52 ms vs. 43.8 ms; 1199.4 ms vs. 1288.8 ms), consistent with run-to-run netem jitter variation rather than an architectural effect.

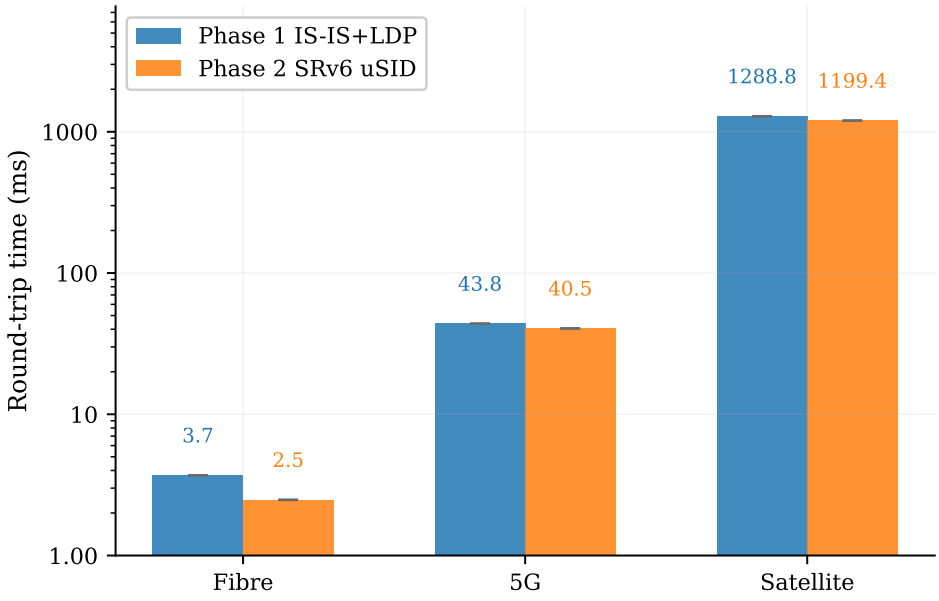


Figure 6.13: Cross-phase WAN bearer RTT comparison, log scale. Fiber, 5G, and satellite bearers for both phases.

6.3.3 Total Delivered Throughput Comparison

Figure 6.14 compares total delivered throughput across all 50 flows in each phase: 10.23 Mbit/s in Phase 1 versus 10.74 Mbit/s in Phase 2. This is approximately 5% higher in Phase 2, with both phases close enough that this is not read as an architectural advantage either way. The gap between the per-flow goodput reported in Sections 6.1.3 and 6.2.3 (hundreds of Mbit/s for mouse flows) and the aggregate figures here is expected: per-flow goodput is computed over each flow's own short active window, whereas the aggregate is a time-average across the full schedule, in which the short-lived mouse flows contribute little and the rate-limited elephant flows dominate.

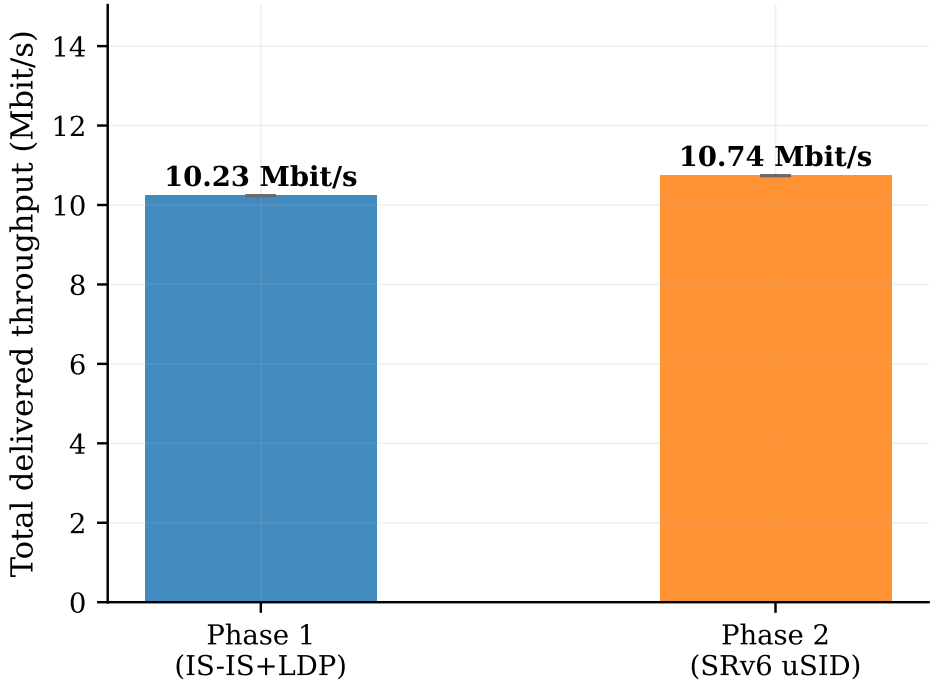


Figure 6.14: Cross-phase total delivered throughput comparison, all 50 flows, both phases.

6.3.4 Failure-Recovery Comparison

A note on statistical scope is appropriate before reading these figures. The steady-state metrics each aggregate up to 500 per-sample probe observations under a fully deterministic, fixed-seed traffic schedule, so within-measurement variance is captured directly in the reported distributions. The Phase 1 IS-IS + LDP reconvergence is a deterministic chain (SPF recomputation followed by LDP label resignalling) with little expected variance at this topology scale, and the IS-IS adjacency telemetry collected in parallel independently corroborates the data-plane observation. The two figures are independently timed runs (200 ms ICMP probing versus 2 s adjacency polling), so their absolute event timestamps differ; it is the adjacency-drop and recovery behaviour, not the wall-clock alignment, that corroborates the data-plane result. The Phase 2 TI-LFA result is measured across 5 failure-injection trials, providing a direct variance estimate for the headline recovery figure. Independent replication across separately deployed sessions is identified as a way to strengthen the steady-state results further (Chapter 7, Section 7.6).

Figure 6.15 places the Phase 1 and Phase 2 failure-recovery timelines side by side in a single composite figure, and Figure 6.16 summarises the comparison as a single log-scale bar chart. Together, these give a visual summary of the most consistent finding in this thesis.

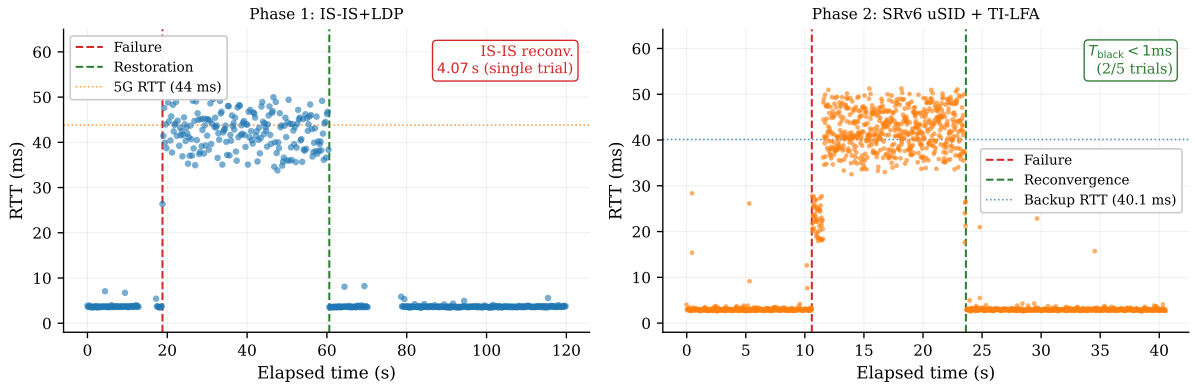


Figure 6.15: Cross-phase failure-recovery timeline (single composite figure). Left: Phase 1 IS-IS + LDP, RTT steps from fiber baseline to the 5G bearer level (44 ms) for 4.07 s following failure injection, returning to baseline on restoration. Right: Phase 2 SRv6 uSID + TI-LFA, $T_{\text{black}} < 1$ ms for 2 of 5 trials, backup-path RTT stable at 40.1 ms. Both panels share the same vertical scale.

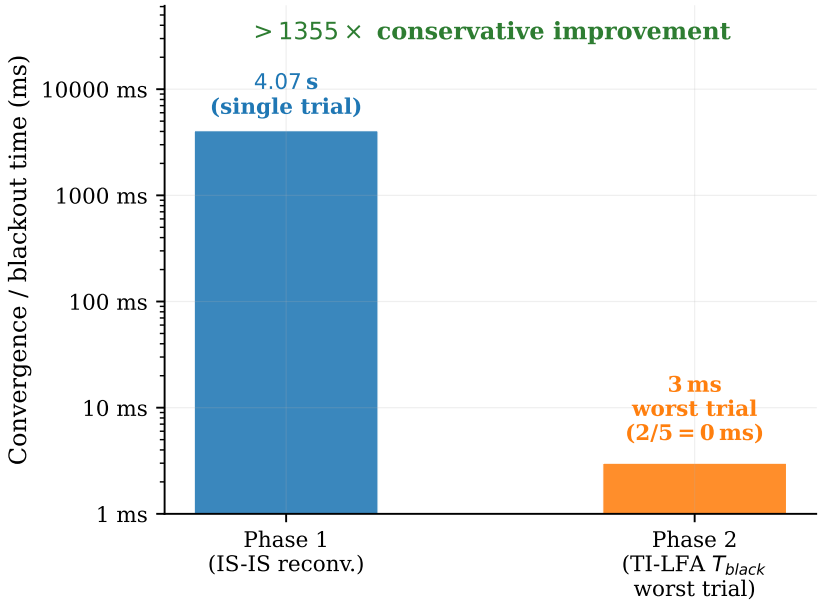


Figure 6.16: Cross-phase convergence/blackout time summary, log scale. Phase 1 IS-IS + LDP reconvergence (4066 ms) versus Phase 2 TI-LFA T_{black} worst trial (3 ms, across 5 trials), giving a conservative improvement exceeding 1355×. Statistical scope is discussed in Section 7.6.

This comparison is the visual counterpart to the numerical result developed in Chapter 7: the worst-trial-based, conservative improvement factor of 1355× is derived from the same two measurements shown here. Its interpretation against the research questions, including the methodological caveats around the differing probe resolutions used in each phase (200 ms in Phase 1, 1 ms in Phase 2), is developed in Chapter 7, Section 7.3.

6.4 DC Fabric Telemetry (Grafana)

Independent of the WAN-side convergence and recovery measurements above, the SR Linux DC-fabric nodes expose IS-IS and interface state via OpenConfig YANG paths, collected by gNMIc and visualised through a Prometheus/Grafana telemetry stack (Chapter 3). This is a separate observability artifact

from the WAN-side netmiko-polled telemetry used to produce the IS-IS telemetry and convergence-summary figures above: where those figures characterise WAN-transport convergence and recovery, the Grafana dashboards characterise the data centre fabric's own interface activity over the same experimental window.

Figures 6.17 and 6.18 show the "Interface Out Octets" panel for each phase's DC fabric, covering all four leaf interfaces across both data centres (DC1edge-leaf, DC1edge-leaf2, DC2edge-leaf, DC2edge-leaf2, plus management interfaces). Both dashboards show clear, correlated bursts of interface activity rather than a flat baseline, confirming that the TrafPy bimodal traffic schedule is genuinely traversing the EVPN-VXLAN data centre fabric on both leaves in both phases. The brief dip between Phase 1's two traffic bursts corresponds to the IS-IS + LDP failure-injection event (Section 6.1.4): DC-fabric interface activity visibly drops while the WAN-side reconvergence is in progress and recovers once the path is restored. Phase 2's pattern of smaller periodic bursts following the main traffic burst corresponds to the 5 individual TI-LFA failure trials (Section 6.2.5), each producing a brief, visible spike as the corresponding short traffic probe is sent during that trial.

Figure 6.17: Phase 1 DC fabric interface activity (Grafana, "Interface Out Octets" panel, 15-minute window). Two traffic bursts are visible, each ramping to approximately 14–15.5 million octets; the dip between them corresponds to the IS-IS + LDP failure-injection event.

Figure 6.18: Phase 2 DC fabric interface activity (Grafana, "Interface Out Octets" panel, 15-minute window). The main traffic burst is followed by 5 smaller periodic bursts, each corresponding to one TI-LFA failure trial.

This is direct, independent evidence for the controlled-variable claim central to this thesis's methodology (Chapter 4): the data centre fabric visibly carries the same traffic schedule in both phases, and its own interface-level activity tracks the WAN-side experimental events (failure injection in Phase 1, individual TI-LFA trials in Phase 2) without itself being reconfigured between phases.

Chapter 7

Discussion

This chapter interprets the Phase 1 and Phase 2 results against the research questions, documents implementation challenges encountered during the build, and evaluates threats to validity.

7.1 Phase 1: Stitched IS-IS + LDP Baseline

The central Phase 1 finding is an IS-IS + LDP convergence window of 4.07 seconds following a fiber WAN access failure. For most DCI workloads, this is operationally disruptive: storage replication jobs time out, VM live migrations stall, and database heartbeats may declare a healthy replica dead. The failure was injected via a clean management-plane interface disable, producing an immediate IS-IS adjacency loss; this represents a relatively favourable case for the stitched architecture rather than an adversarial one, since a silent fiber cut relying on hold-timer expiry would likely take longer to detect before this same LDP re-signalling process even begins.

The Phase 1 fiber RTT figures (Section 6.1) are consistent with each other and with the low latency expected across a four-hop diamond with no contention. IPv6's wider P99 (5.45 ms vs 4.05 ms for the IPv4 backup path) may plausibly reflect additional 6VPE encapsulation processing at the PE routers, though this thesis did not isolate that specific cause with a dedicated measurement and reports the P99 difference as an observation rather than a confirmed mechanism.

The elephant-flow ceiling (0.68 Mbit/s against mouse UDP's 641 Mbit/s) is not unique to Phase 1: Phase 2 measured a closely comparable 0.68 Mbit/s under the same schedule, which rules out the WAN transport as the cause and points to a shared testbed constraint (Section 7.7) rather than a property of either architecture.

7.2 Phase 2: Unified SRv6 uSID Fabric

The central Phase 2 finding is a TI-LFA packet-loss window (T_{black}) of 1.2 ms mean across 5 failure-injection trials, against Phase 1's 4.07-second IS-IS + LDP window, a difference of more than three orders of magnitude. The 5 trials span 0–3 ms, consistent with TI-LFA's pre-computed repair path already being installed at the Point of Local Repair before failure occurs; the small ms-scale variation across trials is consistent with probe-interval quantisation (1 ms probes) rather than necessarily reflecting variability in TI-LFA itself.

The elevated-RTT duration T_{elev} of 13.1 seconds is substantially longer than T_{black} and reflects a structurally different process: T_{black} measures TI-LFA's near-instantaneous local repair, while T_{elev} measures the considerably slower IS-IS SPF re-convergence required to re-advertise the restored fiber path and switch traffic back to the primary. TI-LFA is a fast local-repair mechanism rather than a restoration optimisation, so the long T_{elev} is an expected property of the underlying IS-IS re-convergence process rather than a TI-LFA shortcoming.

The Phase 2 host-to-host fiber RTT mean of 4.13 ms is somewhat higher than Phase 1's 3.70–3.65 ms range, while the router-originated fiber RTT shows the opposite pattern, with Phase 2's 2.477 ms below Phase 1's 3.70 ms. uSID encapsulation processing at the border PEs is a plausible contributor to the host-to-host increase, and simpler single-encapsulation P-router forwarding to the router-originated decrease, but neither was isolated as a confirmed cause, so both are reported as observations rather than explained results. The two measurements probe different points in the forwarding chain (full host-to-host stack

versus router-to-router) and should not be conflated.

7.3 Cross-Phase Comparison

Finding 1: On identical hardware, under the same fiber-link failure and the same traffic schedule, the pre-installed TI-LFA backup in the unified SRv6 uSID fabric bounds the data-plane disruption window (T_{black}) to a mean of 1.2 ms (3 ms at worst across 5 trials), against a 4.07 s IS-IS + LDP reconvergence window in the stitched baseline: a conservative worst-trial improvement exceeding 1355×.

Finding 2: Steady-state performance and per-packet overhead show no comparable advantage for the unified transport. Throughput is statistically indistinguishable between phases, steady-state latency is at best neutral, and the WAN-leg encapsulation is 32 bytes heavier for SRv6 uSID than for the MPLS label stack at this single-hop scale, the expected direction since uSID compression is a multi-hop property that a single-hop topology cannot exhibit.

The recovery comparison in Figure 6.16 deliberately uses Phase 2's worst trial rather than its mean, giving a conservative lower bound rather than an optimistic headline figure. Steady-state latency tells a more mixed story: depending on whether the host-to-host or router-originated measurement is used, Phase 2 shows either a modest RTT increase or a modest decrease relative to Phase 1, with no confirmed mechanism for either direction (Section 7.2); throughput is statistically indistinguishable, since both phases share the same DC fabric and the same testbed-level MTU constraint affecting elephant flows. The clearest architectural difference between the two phases is therefore failure-recovery time, not steady-state performance or per-packet overhead, consistent with TI-LFA being a fast-reroute mechanism rather than a general performance optimisation.

7.4 Answers to the Research Questions

Finding 3: The two halves of the unified architecture differ in deployment maturity. Flex-Algo delay-metric path differentiation, the distributed per-node mechanism, is deployable and verifiable today; the centralised control plane was taken as far as established, synchronised PCEP and BGP-LS sessions on a multi-vendor software combination, with end-to-end PCE-initiated SRv6 path steering remaining as the next stage.

RQ1. Phase 2 shows that WAN transport unification under a single SRv6 uSID domain is achievable on Nokia SR OS 25.7.R2: End.uDT4 provides the VPN/tenant-forwarding function at the PE, uN provides node-level transit forwarding at the P-routers, and Flex-Algo (Algo 0 / Algo 128) provides metric-based path differentiation, all confirmed operational via direct SID-database verification. This finding needs to be stated precisely: the DC fabric (EVPN-VXLAN, leaf-spine) is deliberately held constant across both phases as a controlled experimental variable, so the unification demonstrated here is specifically WAN-transport unification, not end-to-end elimination of every protocol boundary in the architecture. The DC-fabric-to-WAN handoff at the border PE remains a structural boundary in both phases.
RQ2. The TI-LFA results provide a clear, quantitative answer to this question for the single failure scenario tested. The unified fabric showed substantially faster failure recovery by activating a pre-computed repair path at the Point of Local Repair, avoiding the SPF-recomputation-then-LDP-resignalling sequence that dominates Phase 1's recovery time. This finding is scoped to the fiber-bearer, single-link failure scenario evaluated here (Section 7.6), and should not be read as a general claim about resilience across all failure types or bearer types.
RQ3. Flex-Algo 0 and Flex-Algo 128 are confirmed operational, which shows that delay-metric-based path differentiation works as designed without needing a PCE at all. On the centralised side, both PCEP and BGP-LS sessions were brought up between the border PEs and the OpenDaylight 0.18.2 PCE. PCEP reached full synchronisation on both borders, with stateful and PCE-initiated LSP capabilities negotiated. BGP-LS sessions reached Established state with the LINKSTATE address family, and the ODL BGP RIB was instantiated with both peers registered as eBGP. The PCEP topology provider's

TED reference was configured to point to the BGP-LS linkstate topology, so the two protocols are properly connected on the ODL side. Taken together, this establishes a working centralised control-plane foundation: the sessions, capabilities, and topology wiring that PCE-driven path computation depends on are all in place and verified. This thesis scopes RQ3 to establishing and validating that foundation, alongside the Flex-Algo result; end-to-end PCE-initiated SRv6 path instantiation is the subsequent stage that builds directly on it, and is set out as future work in Chapter 8.

7.5 Implementation Challenges

Several testbed-specific constraints are documented here for reproducibility, since none were found in existing Containerlab or Nokia SR-SIM literature before this work.

`tc netem` does not affect Nokia SR OS containers: SR OS bypasses the Linux kernel networking stack entirely for data-plane forwarding, so failure injection must instead use `admin-state disable` on the relevant IS-IS interface via the SR OS management plane (netmiko over SSH), not kernel-level interface manipulation.

`docker exec` commands targeting the SR OS or SR Linux CLI directly do not reliably reflect data-plane state: SR OS containers process management-plane commands in a separate context from the data plane, and ping/connectivity tests issued without explicitly specifying the correct routing or network instance (e.g. `router-instance Base` on SR OS, `network-instance default` on SR Linux) silently test the management plane instead, producing misleading results that appear as connectivity failures despite the actual data plane functioning correctly. All measurement automation in this work explicitly targets the correct instance for this reason.

Phase 1's EVPN-VXLAN DC fabric requires a second leaf per data centre to genuinely exercise VXLAN tunnelling between leaves; a single-leaf topology allows traffic between hosts to be locally bridged without ever crossing a tunnel, which would not constitute a valid test of the EVPN-VXLAN mechanism. The second leaf, `leaf2`, required an `advertise-neighbor-type router-host` setting not present on `leaf1`, since `leaf1`'s standard IRB configuration is incompatible with that setting; this asymmetric configuration between the two leaves was a deliberate, documented choice made to achieve genuine multi-leaf EVPN-VXLAN bridging, not an oversight, and is described at the configuration level in Chapter 5.

The two phases differ in the address families carried over the WAN-transport leg, and this is a scope decision rather than a property of either architecture. On the Phase 1 stitched IS-IS + LDP transport, the dual-stack 6VPE service [29] carries cross-DC tenant traffic over both address families, and the IPv6 primary path is the one used for the headline Phase 1 fiber RTT measurement (Chapter 6, Section 6.1). Phase 2 was deployed with the IPv4 tenant service (End.uDT4) only over the unified SRv6 path; the VPN-IPv6 address family was not brought up on the border-to-border session in this iteration, and the Phase 2 DC fabric was addressed for IPv4 only. Cross-DC IPv6 forwarding over the unified SRv6 path (for example via an End.uDT6 service) is therefore out of scope for the measurements reported here and is identified as future work (Chapter 8). Because the per-packet encapsulation comparison in this thesis uses an identical IPv4 inner payload in both phases (Section 6.2.4), the address-family scope of the Phase 2 service does not affect that comparison; it affects only the steady-state RTT figures, where the small in-phase IPv4/IPv6 difference is discussed in Section 7.1.

7.6 Threats to Validity

External validity. SR-SIM executes the same SR OS control-plane software as physical Nokia hardware, so convergence timing and protocol behaviour are expected to be reasonably faithful to the physical platform. Software-based data-plane forwarding does, however, limit achievable throughput, meaning the goodput figures reported here are best read as relative indicators between phases rather than absolute capacity claims. The two-DC testbed topology (Chapter 5) is smaller than a production deployment; IS-IS + LDP convergence time would plausibly be longer at production scale, which, if anything, would make the gap between the two architectures' recovery times more pronounced rather than less, though this was not tested directly. Only fiber-bearer failures were evaluated in this work; 5G and satellite link failures are left for future work. The Grafana-visualised DC-fabric telemetry (Chapter 6, Section 6.4) supports the controlled-variable assumption underlying this comparison: the DC-fabric interface activity visibly carries the same traffic schedule in both phases, dips during Phase 1's failure-injection event, and shows a distinct small burst per TI-LFA trial in Phase 2, consistent with the data centre fabric's own behaviour tracking the WAN-side experimental events without itself being reconfigured between phases.

This is offered as supporting evidence for attributing the measured differences in Sections 7.1–7.3 to the WAN-transport architecture, rather than as definitive proof ruling out every possible confound.

A further external-validity concern is vendor generalisability. The specific timing numbers in this thesis (IS-IS + LDP convergence at 4.07 s, TI-LFA T_{black} at 1.2 ms mean) are measurements on Nokia SR OS 25.7.R2, and may differ on other implementations. The qualitative finding, however, is grounded in the protocol standards themselves: the IS-IS + LDP reconvergence chain (SPF recomputation followed by LDP label resignalling) and TI-LFA's pre-installed backup path activation are both RFC-defined behaviours, and the architectural reason why TI-LFA avoids the reconvergence chain applies equally to any conformant implementation, such as Cisco IOS XR or Juniper Junos, both of which document TI-LFA support for segment routing. On hardware ASIC platforms, both phases would likely show faster absolute timing, but the relative advantage of skipping SPF and LDP resignalling entirely should remain. The address-family scope of the Phase 2 unified-path service (IPv4-only, Section 7.5) is a property of this particular deployment iteration and should not be read as a constraint of SRv6 or of either architecture. Whether the qualitative convergence advantage and the encapsulation scaling behaviour hold across other vendor implementations is left as future work.

Construct validity. ICMP probes measure data-plane reachability and may not fully represent every class of application-level recovery behaviour, though the two are closely coupled in both architectures tested. The TrafPy exponential inter-arrival model is a simplification of real data centre traffic burstiness and temporal correlation.

Conclusion validity. The Phase 2 TI-LFA recovery figure is measured across 5 failure-injection trials, giving a direct variance estimate (1.2 ms $\pm$ 1.2 ms); a larger number of independent deployments would further tighten this estimate. The Phase 1 convergence figure is supported by IS-IS adjacency telemetry corroborating the data-plane observation, and independent replication would strengthen it further. The steady-state metrics each aggregate hundreds of per-sample probes under a deterministic schedule, so their within-measurement variance is captured directly.

7.7 Limitations

The principal limitations of this work, each discussed in context above, are:

- **Software data plane.** Absolute throughput reflects the SR-SIM forwarding ceiling rather than physical Nokia hardware capacity and is valid only as a relative indicator between phases.
- **PCE scope.** The centralised evaluation is bounded by established, validated PCEP and BGP-LS sessions; end-to-end PCE-initiated SRv6 instantiation is left for follow-on work, so RQ3 is answered with respect to Flex-Algo path differentiation and control-plane session establishment.
- **Elephant-flow ceiling.** Shown to be testbed-wide rather than architecture-specific, but not root-caused beyond a plausible MTU/MSS interaction.
- **Single-bearer failures only.** Simultaneous multi-bearer degradation is left for future work.

Chapter 8

Conclusion

8.1 Summary of Findings

This thesis set out to measure, rather than argue for, the effect of replacing a stitched IS-IS + LDP WAN transport with a unified SRv6 uSID WAN transport in a distributed data centre interconnect, holding the data centre fabric itself constant. Three findings stand out.

First, failure-recovery time improves substantially (Finding 1): more than three orders of magnitude, on the same hardware, the same fiber failure, and the same traffic schedule, with a conservative worst-trial-based improvement factor of 1355×. This is the clearest and most defensible result in this thesis.

Second, steady-state latency and DCN throughput show no comparable advantage for the unified transport, and at this testbed's single-hop scale the WAN-leg encapsulation is heavier, not lighter (Finding 2): 32 bytes more for SRv6 uSID than for the MPLS label stack (138 B versus 106 B on the wire). This is consistent with uSID's compression being a multi-hop property, so the single-hop topology evaluated here cannot exhibit it; a larger WAN topology, identified in the future-work section, would measure it directly.

Third, the centralised-control vision most often associated with SR-based WAN unification was taken in this thesis as far as a working, validated control-plane foundation on a multi-vendor software combination (Finding 3). Both PCEP and BGP-LS sessions were brought up between the border routers and the OpenDaylight PCE: PCEP reached full synchronisation with PCE-initiated LSP capability negotiation on both borders, and BGP-LS sessions reached Established state with the LINKSTATE address family. End-to-end PCE-initiated SRv6 path instantiation builds on this foundation and is scoped as the next stage of the work. Flex-Algo-based steering, which operates independently of the PCE, is the distributed, per-node path-differentiation element that this thesis demonstrates end-to-end.

8.2 Contributions Revisited

The four contributions stated in Chapter 1 are assessed here against what was actually achieved.

Contribution 1 (IS-IS + LDP convergence measurement on Nokia SR OS) was delivered as stated. The 4.07 s reconvergence window, with IS-IS adjacency telemetry providing independent corroboration of the data-plane observation, constitutes the stitched baseline. As noted throughout, this is a measurement on a virtualised platform rather than physical hardware, a point stated in Section 7.6.

Contribution 2 (direct TI-LFA vs IS-IS + LDP convergence comparison) was delivered. The 1355× conservative improvement factor is derived from Phase 2's worst-trial T_{black} (3 ms, across 5 failure-injection trials) against Phase 1's baseline (4066 ms).

Contribution 3 (Flex-Algo path steering and PCE deployment evaluation) was delivered as scoped. Flex-Algo path differentiation across two independent forwarding topologies (Algo 0 and Algo 128) was confirmed operational. On the PCE side, both PCEP and BGP-LS sessions were brought to Established state, with PCEP fully synchronised and the LINKSTATE address family negotiated on BGP-LS, giving a working centralised control-plane foundation. End-to-end PCE-initiated SRv6 path instantiation builds on this foundation and is set out as future work.

Contribution 4 (reproducible Containerlab testbed and implementation observations) was delivered. Both topologies are version-controlled and deployable from a single command on any host running Nokia SR-SIM. The two implementation observations (the single-leaf EVPN-VXLAN topology requirement

and the IPv4-only address-family scope of the Phase 2 unified-path service) are documented at a level of specificity sufficient for independent reproduction or avoidance.

8.3 Future Work

Four directions follow most directly from this thesis's stated limitations. First, repeating the TI-LFA and convergence measurements across multiple independent deployments would test whether the recovery-time advantage holds consistently and allow even stronger statistical claims; this would also enable a fully symmetric comparison (equal number of failure injections in both phases). Second, extending the WAN topology beyond four routers, even modestly to six or eight, would allow uSID's compression advantage to be measured directly rather than only argued for theoretically, closing the gap this thesis explicitly leaves open. Third, deploying TWAMP-based dynamic delay measurement would replace the static IS-IS TE delay attributes currently used by Flex-Algo 128 with real-time measured values, enabling evaluation under realistic jitter and delay variation. Fourth, exercising end-to-end PCE-initiated SRv6 path instantiation would build directly on the PCEP and BGP-LS session establishment demonstrated in this work, completing the centralised path-steering side of RQ3.

8.4 Closing Remark

The most useful outcome of this thesis may not be the headline improvement factor, substantial as it is, but the discipline of measuring rather than assuming: several claims that circulate informally about unified SR fabrics, uniform efficiency gains, straightforward centralised control, encapsulation overhead that is simply "smaller", turn out, on direct measurement, to be true only within specific scope conditions that are easy to state once measured and easy to overstate if not. This thesis's contribution is as much in stating those scope conditions precisely as in the headline number itself.

Bibliography

[1] S. Troia, L. Borgianni, G. Sguotti, S. Giordano, and G. Maier, "A Comprehensive Survey on Software-Defined Wide Area Network," *IEEE Communications Surveys & Tutorials*, vol. 28, pp. 2805–2845, 2026. DOI: 10.1109/COMST.2025.3594678.

[2] Y. Gan et al., "An open-source benchmark suite for microservices and their hardware-software implications for cloud & edge systems," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'19)*, ACM, 2019. DOI: 10.1145/3297858.3304013.

[3] D. Narayanan et al., "Efficient large-scale language model training on GPU clusters using Megatron-LM," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'21)*, ACM, 2021. DOI: 10.48550/arXiv.2104.04473.

[4] Y. Akharchaf and G. Gao, "Enhancing Traffic Engineering with AI: Comparative Analysis of MPLS, SD-WAN, and SRv6," *International Journal of Research and Innovation in Social Science*, vol. 9, no. 11, Nov. 2025. DOI: 10.47772/IJRISS.2025.91100027.

[5] J. Drake et al., *BGP MPLS-Based Ethernet VPN*, RFC 7432, Feb. 2015. DOI: 10.17487/RFC7432.

[6] C. Filsfils, P. Camarillo, J. Leddy, D. Voyer, S. Matsushima, and Z. Li, *Segment Routing over IPv6 (SRv6) Network Programming*, RFC 8986, Feb. 2021. DOI: 10.17487/RFC8986.

[7] A. Tulumello et al., "Micro SIDs: A Solution for Efficient Representation of Segment IDs in SRv6 Networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 774–786, Mar. 2023. DOI: 10.1109/TNSM.2022.3205265.

[8] A. Farrel, J.-P. Vasseur, and J. Ash, *A Path Computation Element (PCE)-Based Architecture*, RFC 4655, Aug. 2006. DOI: 10.17487/RFC4655.

[9] E. Crabbe, I. Minei, J. Medved, and R. Varga, *Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE*, RFC 8231, Sep. 2017. DOI: 10.17487/RFC8231.

[10] P. Psenak, S. Hegde, C. Filsfils, K. Talaulikar, and A. Gulko, *IGP Flexible Algorithm*, RFC 9350, Feb. 2023. DOI: 10.17487/RFC9350.

[11] S. Deering and R. Hinden, *Internet Protocol, Version 6 (IPv6) Specification*, RFC 8200, Jul. 2017. DOI: 10.17487/RFC8200.

[12] J. Rabadan, S. Sathappan, W. Henderickx, A. Sajassi, and J. Drake, *Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks*, RFC 9014, May 2021. DOI: 10.17487/RFC9014.

[13] A. Bashandy, S. Litkowski, C. Filsfils, P. Francois, B. Decraene, and D. Voyer, *Topology Independent Fast Reroute Using Segment Routing*, RFC 9855, Oct. 2025. DOI: 10.17487/RFC9855.

[14] L. Ginsberg, S. Previdi, S. Giacalone, D. Ward, J. Drake, and Q. Wu, *IS-IS Traffic Engineering (TE) Metric Extensions*, RFC 8570, Mar. 2019. DOI: 10.17487/RFC8570.

[15] M. Srndic and R. Docter, *Containerlab: Lightweight Network Emulation with Containers*, FOSDEM 2022, Brussels, Belgium, Available: https://containerlab.dev, Feb. 2022.

[16] S. Sezer et al., "Are We Ready for SDN? Implementation Challenges for Software-Defined Networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, 2013. DOI: 10.1109/MCOM.2013.6553676.

[17] S. Jain et al., "B4: Experience with a Globally-Deployed Software Defined WAN," in *Proceedings of the ACM SIGCOMM Conference*, 2013, pp. 3–14. DOI: 10.1145/2486001.2486019.

[18] C.-Y. Hong et al., "Achieving High Utilization with Software-Driven WAN," in *Proceedings of the ACM SIGCOMM Conference*, 2013, pp. 15–26. DOI: 10.1145/2486001.2486012.

[19] P. François, C. Filsfils, J. Evans, and O. Bonaventure, "Achieving Sub-Second IGP Convergence in Large IP Networks," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 3, pp. 33–44, 2005. DOI: 10.1145/1070873.1070877.

[20] C. Filsfils, S. Previdi, L. Ginsberg, B. Decraene, S. Litkowski, and R. Shakir, *Segment Routing Architecture*, RFC 8402, Jul. 2018. DOI: 10.17487/RFC8402.

[21] P. L. Ventre et al., "Segment Routing: A Comprehensive Survey of Research Activities, Standardization Efforts, and Implementation Results," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 182–221, 2021. DOI: 10.1109/COMST.2020.3036826.

[22] A. Abdelsalam et al., "SRPerf: A Performance Evaluation Framework for IPv6 Segment Routing," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 2320–2333, 2021. DOI: 10.1109/TNSM.2020.3048328.

[23] C. Scarpitta, G. Sidoretti, A. Mayer, S. Salsano, A. Abdelsalam, and C. Filsfils, "High Performance Delay Monitoring for SRv6-Based SD-WANs," *IEEE Transactions on Network and Service Management*, vol. 21, no. 1, pp. 1067–1080, 2024. DOI: 10.1109/TNSM.2023.3300151.

[24] J. Abhishek Singh, M. R. Sachin Kumar, and K. S. Shushrutha, "Implementation of Topology Independent Loop Free Alternate with Segment Routing Traffic," in *Proceedings of the 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021. DOI: 10.1109/ICCCNT51525.2021.9579530.

[25] L. Roelens, Ó. González de Dios, I. de Miguel, E. Echeverry, and R. J. Durán Barroso, "Performance Evaluation of TI-LFA in Traffic-Engineered Segment Routing-Based Networks," in *Proceedings of the 19th International Conference on Design of Reliable Communication Networks (DRCN)*, 2023. DOI: 10.1109/DRCN57075.2023.10108280.

[26] Y. Rekhter and E. C. Rosen, *BGP/MPLS IP Virtual Private Networks (VPNs)*, RFC 4364, Feb. 2006. DOI: 10.17487/RFC4364.

[27] M. Mahalingam et al., *Virtual eXtensible Local Area Network (VXLAN)*, RFC 7348, Aug. 2014. DOI: 10.17487/RFC7348.

[28] D. Katz and D. Ward, *Bidirectional Forwarding Detection (BFD)*, RFC 5880, Jun. 2010. DOI: 10.17487/RFC5880.

[29] J. D. Clercq, D. Ooms, M. Carugi, and F. L. Faucheur, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*, RFC 4659, Sep. 2006. DOI: 10.17487/RFC4659.

[30] S. Litkowski, S. Agrawal, K. Ananthamurthy, and K. Patel, *Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop*, RFC 8950, Nov. 2020. DOI: 10.17487/RFC8950.

[31] C. Filsfils, D. Dukes, S. Previdi, J. Leddy, S. Matsushima, and D. Voyer, *IPv6 Segment Routing Header (SRH)*, RFC 8754, Mar. 2020. DOI: 10.17487/RFC8754.

[32] R. Callon, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*, RFC 1195, Dec. 1990. DOI: 10.17487/RFC1195.

[33] H. Gredler, J. Medved, S. Previdi, A. Farrel, and S. Ray, *Distribution of Link-State and Traffic Engineering Information Using BGP*, RFC 9552, Nov. 2023. DOI: 10.17487/RFC9552.

[34] Nokia, *SR OS Simulator (SR-SIM): Containerized Network OS for Lab and Testing*, Accessed: May 2026, Nokia, 2025. [Online]. Available: https://containerlab.dev/manual/kinds/sros/.

[35] C. W. Parsonson, J. L. Benjamin, and G. Zervas, "Traffic Generation for Benchmarking Data Centre Networks," *Optical Switching and Networking*, vol. 46, p. 100 695, 2022. DOI: 10.1016/j.osn.2022.100695.

[36] OpenConfig Working Group, *OpenConfig: Vendor-Neutral Data Models for Network Devices*, https://openconfig.net, Accessed: May 2026, 2024.

[37] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis, *Framework for IP Performance Metrics*, RFC 2330, May 1998. DOI: 10.17487/RFC2330.

[38] B. Constantine, G. Forget, R. Geib, and R. Schrage, *Framework for TCP Throughput Testing*, RFC 6349, Aug. 2011. DOI: 10.17487/RFC6349.

[39] G. Almes, S. Kalidindi, and M. Zekauskas, *A Round-trip Delay Metric for IPPM*, RFC 2681, Sep. 1999. DOI: 10.17487/RFC2681.

[40] S. Hemminger, *Network Emulation with NetEm*, Linux Conf Australia, Canberra, Australia, Apr. 2005. [Online]. Available: https://www.rationali.st/blog/files/20151126-jittertrap/netem-shemminger.pdf.

[41] ITU-R, "Minimum requirements related to technical performance for imt-2020 radio interface(s)," International Telecommunication Union, Tech. Rep. Report ITU-R M.2410-0, 2017. [Online]. Available: https://www.itu.int/pub/R-REP-M.2410.

[42] 3GPP, "Service requirements for the 5g system; stage 1 (release 16)," 3rd Generation Partnership Project, Tech. Rep. Technical Specification (TS) 22.261, 2020. [Online]. Available: https://www.3gpp.org/specifications-technologies/specs-by-series/22-series.

[43] M. Allman, D. Glover, and L. Sanchez, *Enhancing TCP Over Satellite Channels using Standard Mechanisms*, RFC 2488, 1999. DOI: 10.17487/RFC2488.

[44] Nokia, *SR Linux (SRL): Containerized Network OS for Lab and Testing*, Containerlab, Accessed: Jun. 2026, 2025. [Online]. Available: https://containerlab.dev/manual/kinds/srl/.

[45] E. Rosen, A. Viswanathan, and R. Callon, *Multiprotocol Label Switching Architecture*, RFC 3031, Jan. 2001. DOI: 10.17487/RFC3031.