Accepted for publication on July 26th 2025 in IEEE Communications Surveys & Tutorials

A Comprehensive Survey on Software-Defined Wide Area Network

Sebastian Troia, Luca Borgianni, Giacomo Sguotti, Stefano Giordano and Guido Maier

***Abstract*—The rise of Software-Defined Wide Area Network (SD-WAN) a decade ago has ushered in a new era of networking opportunities for organizations worldwide. As enterprises undergo digital transformation fueled by cloud computing adoption and the imperative for enhanced connectivity, SD-WAN has emerged as a transformative paradigm. This survey paper unpacks the intricate ecosystem of SD-WAN technology, offering a deep dive into its foundational principles, architectural components, and diverse deployment scenarios. We explore modern SD-WAN architectures, analyzing key elements such as Customer Premises Equipment (CPE), underlay and overlay networks, controllers, and orchestration planes. By shedding light on the mechanisms that drive SD-WAN's agility and efficiency, we provide critical insights into its transformative potential. Beyond architecture, we examine essential operational aspects, including traffic engineering, performance monitoring, and network resources management within the SD-WAN framework. We also identify key challenges and research gaps, charting a path for future advancements in the field. This comprehensive study serves as a valuable resource for academics, researchers, and industry professionals navigating the evolving landscape of enterprise networking in the era of cloud-driven digital transformation.**

***Index Terms*—Software-Defined Wide Area Network (SD-WAN), Software-Defined Networking (SDN), Digital Transformation, Enterprise Network, Customer Premises Equipment (CPE), Multi-protocol Label Switching (MPLS), Traffic Engineering (TE), Monitoring.**

I. Introduction

Software-Defined Wide Area Network (SD-WAN) has emerged as a revolutionary technology, addressing the limitations of traditional Wide Area Network (WAN) architectures by integrating Software-Defined Networking (SDN) principles into Enterprise networking. First appearing around 2014 [1], the term was coined by Gartner to create a commercial product category that defines how the Enterprise can purchase SDN services from service providers and vendors. SD-WAN rapidly gained traction, not only as a buzzword in industry circles but also as a critical enabler of modern, flexible, and cost-efficient network infrastructures. It quickly evolved into a technical and operational paradigm, underpinned by its ability to decouple network control from the underlying hardware and deliver centralized, software-driven management.

Traditional WANs, primarily based on technologies like Asynchronous Transfer Mode (ATM) [2], Frame Relay [3] and Multi-protocol Label Switching (MPLS) [4], have historically delivered reliable connectivity with Quality of Service (QoS) guarantees. However, these systems often fall short in terms of flexibility, cost-efficiency, and scalability, particularly in the face of evolving business demands [5].

Let us consider a practical example of a company evolving from a start-up to a large enterprise in just a few years. Initially, the company's internal (and private) network, or Intranet, needs to connect the headquarters with factories located in nearby towns. Thus, a single regional WAN carrier can be contracted. Currently, a widely adopted technology for providing QoS-guaranteed WAN connections, particularly for enterprise customers, is MPLS in its commercial product form. Therefore, the company is provided with MPLS regional connectivity after a certain deployment time. Then, the company starts to expand on a national scale and acquires branch sites in the capital city and some of the most important cities. Thus, besides the regional, a national WAN provider has to be contracted. As the company starts to open sales points across the country, perhaps other regional WANs are needed. Then, the company grows to an international and then global enterprise size, with various acquisitions of other brands and operating sales points in many countries. Imagine the interconnection of a branch office on another continent with MPLS: it needs the involvement of an intercontinental carrier to reach the foreign country, and then a national carrier and perhaps also a local regional carrier, each contracted separately for MPLS connectivity. This is repeated thousands of times for each remote branch office.

A solution in which the control is shifted to the edge of the network, close to the company premises, and the universal Internet connectivity is exploited to cross the geographical distance would surely be a great relief: this is what SD-WAN technology can provide. In SD-WAN, SDN is adopted to make the nodes interfacing the enterprise's premises to the WAN programmable and decision-makers, thus creating an overlay network over the Internet by means of different underlay access networks technologies (4G/LTE, xDSL, etc.).

A hallmark of SD-WAN is its ability to construct overlay networks atop existing underlay infrastructures. This abstraction enables enterprises to implement sophisticated Traffic Engineering (TE) techniques, dynamic failover mechanisms, and end-to-end encryption, all while maintaining consistent performance and policy compliance. At its core, SD-WAN integrates the control and management planes into a centralized framework, simplifying network operations and enabling real-time visibility.

This survey delves into SD-WAN's architecture, deployment scenarios, and technological underpinnings. It begins with a comprehensive analysis of SD-WAN's core components,

Sebastian Troia, Guido Maier and Giacomo Sguotti are with the Department of Electronic, Information and Bioengineering (DEIB), Politecnico di Milano, Milan, Italy. Corresponding author e-mail: sebastian.troia@polimi.it.
Luca Borgianni and Stefano Giordano are with the Department of Information Engineering, University of Pisa, Pisa, Italy.

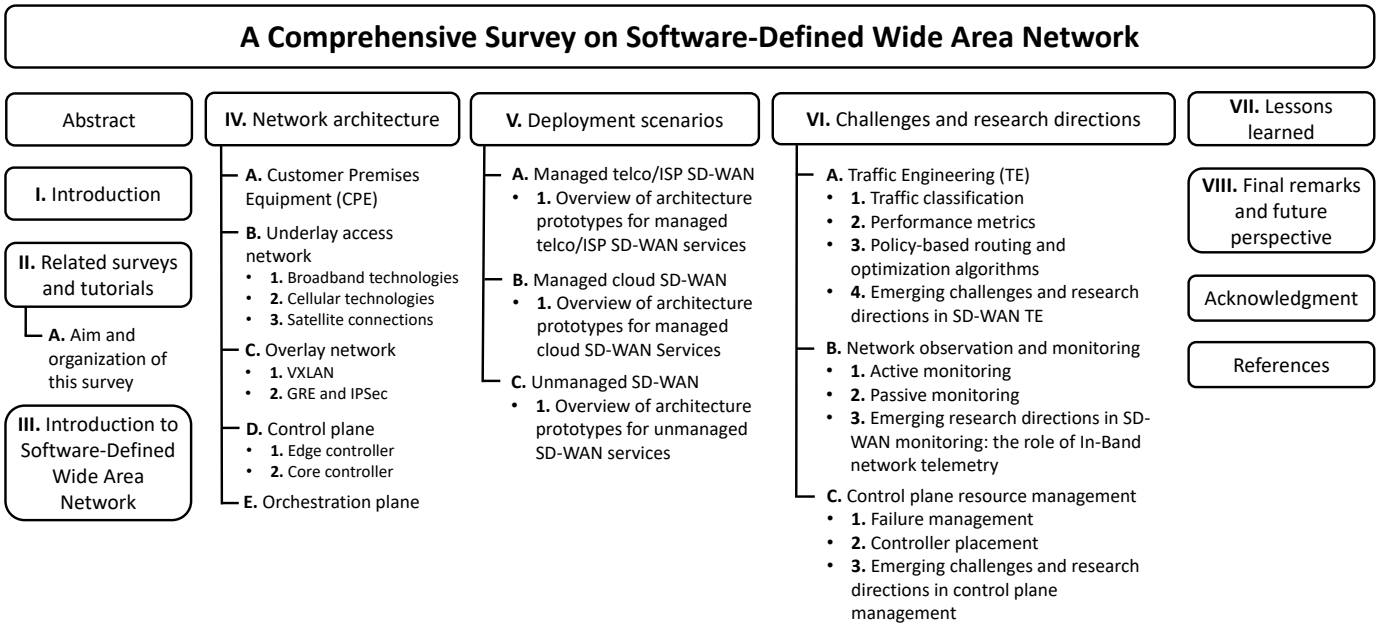

Fig. 1: Table of content of this survey.

including Customer Premises Equipment (CPE), SD-WAN controllers, and overlay/underlay networks. The unique requirements of SD-WAN architectures, such as their need for an agile control plane and adaptive routing capabilities, are discussed in depth. Emerging research challenges and opportunities in SD-WAN are also a focal point of this survey. These include the integration of Artificial Intelligence (AI) and Machine Learning (ML) for predictive traffic management, the adoption of different control plane architectures to enhance resilience, and the use of in-band telemetry for fine-grained network monitoring.

SD-WAN's significance is deeply practical, as evidenced by its widespread adoption across various industries. Many companies have implemented SD-WAN solutions to enhance network performance, security, and flexibility. For instance, in the retail sector, SD-WAN enables rapid deployment of secure connectivity across multiple locations, streamlining operations and improving customer experiences. One of the earliest SD-WAN adopters was GAP Inc., a world-famous clothes retailer. The agreement with the Viptela SD-WAN provider (now acquired by CISCO) to connect GAP San Francisco headquarters to the numerous shops was reported even by the non-technical press [6]. The company reported being able to connect up to 25 shops per night, incrementing connectivity bandwidth by 10-15 times. They planned to connect 1200 selling points, forecasting savings of 50%.

Another case of SD-WAN adoption we would like to mention is reported, e.g., in [7]. In the 2018 Japan Formula-1 Grand Prix, the McLaren team decided for the first time to adopt SD-WAN (solution provided by NTT) to connect their headquarters (McLaren Technology Centre in Woking, UK) to the facilities of the Suzuka circuit. In this way, racing telemetry data gathered by McLaren cars running on the circuit could be processed in the cloud, so real-time indication could be provided back at the race site with a very short round-trip delay. The solution tested in that Grand Prix was to be adopted in all the subsequent F1 events. This example is interesting because F1 events change location each time: only the agility and short deployment time provided by SD-WAN are able to allow setting up an intercontinental high-performance Intranet with an always-new topology and site location every one or two weeks.

SD-WAN's adaptability to diverse organizational needs has made it an essential solution across multiple sectors. In education [8], SD-WAN enables secure and scalable connectivity for K-12[1] school districts and higher education institutions [9], addressing challenges such as remote learning and multi-domain integration. The healthcare sector benefits from SD-WAN's ability to support telehealth and wearable expert systems [10] [11], enabling real-time monitoring and remote patient interactions. In finance, SD-WAN enhances network resilience and operational efficiency, as demonstrated by banking institutions integrating SD-WAN for disaster recovery and load balancing [12]. Similarly, the energy sector leverages SD-WAN to optimize resource allocation and support ultra-low latency applications like electric vehicle charging stations [13]. In transportation, SD-WAN aids in the integration of 5G technologies and network slicing within vehicular networks [14], while in environmental control, it underpins advanced edge-computing architectures for AI-driven applications in smart agriculture and wildlife protection [15]. To conclude, this survey integrates insights from both academic research and industry practices, offering a comprehensive perspective on SD-WAN's current state and future potential. By analyzing its architecture, deployment strategies, and technological advancements, we aim to provide a resource that guides researchers, practitioners, and decision-makers in navigating the complexities of SD-WAN and unlocking its full potential. To enhance readability and accessibility, we include a Table

[1]The expression "K–12" is a shortening of kindergarten (K) for 5–6 year olds through twelfth grade (12) for 17–18 year-olds in the United States and Canada.

TABLE I: Glossary of acronyms.

Acronym	Description	Acronym	Description	Acronym	Description	Acronym	Description
AI	Artificial Intelligence	ISP	Internet Service Provider	OSPF	Open Short Path First	TCP	Transmission Control Protocol
API	Application Program Interface	IPsec	IP security	SCM	SD-WAN Controller Monitor	TE	Traffic Engineering
BFD	Bidirectional Forwarding Detection	LAN	Local Area Network	QoE	Quality of Experience	uCPE	universal CPE
BGP	Border Gateway Protocol	MARL	Multi-Agent Reinforcement Learning	QoS	Quality of Service	UDP	User Datagram Protocol
CPE	Customer Premises Equipment	MEF	Metro Ethernet Forum	RL	Reinforcement Learning	USP	Unmanaged SD-WAN Service Provider
CSP	Cloud Service Provider	ML	Machine Learning	RTT	Round Trip Time	vCPE	virtual CPE
DC	Data Center	MPLS	Multi-protocol Label Switching	SCM	SD-WAN Controller Monitor	VLAN	Virtual LAN
DoS	Denial of Service	MPTCP	Multi-Path TCP	SD-WAN	Software-Defined Wide Area Network	VM	Virtual Machine
DQN	Deep Q-Network	MSP	Managed SD-WAN Service Provider	SDN	Software-Defined Networking	VNI	Virtual Network Identifier
DRL	Deep Reinforcement Learning	NBI	Network-Based Interface	SFC	Service Function Chain	VoIP	Voice over IP
GRE	Generic Routing Encapsulation	Net-KPIs	Network Key Performance Indicators	SONET	Synchronous Optical NETworking	VTEP	VXLAN Tunnel End Point
ICMP	Internet Control Message Protocol	NS	Network Service	SRv6	Segment Routing over IPv6	VXLAN	Virtual eXtensible LAN
INT	In-band Network Telemetry	ONOS	Open Network Operating System	STAMP	Simple Two-way Active Measurement Protocol	WAN	Wide Area Network

of Contents on page 2 outlining the structure of the survey, as well as a glossary of acronyms for reference in Table I[2].

II. Related Surveys and Tutorials

The landscape of SDN has been explored extensively through numerous survey papers, each highlighting different facets of this dynamic field. In this section, we have organized the related work into two distinct groups. The first group includes surveys that discuss SDN and its relevance to SD-WAN, addressing SD-WAN as a topic only in passing or as an extension of SDN concepts. The second group focuses on surveys that directly consider SD-WAN as their primary topic. Here, we analyze these works, highlighting their contributions while positioning our efforts as a step toward bridging the existing gaps and providing a more comprehensive and focused survey on SD-WAN technology.

Previous works have offered broad perspectives on SDN technologies spanning diverse domains, including DCs, wireless networks, optical networks, and transport networks [16], [17], [18], [19].

The survey presented in [16] provides a thorough overview of the SDN paradigm, with a particular focus on OpenFlow (as also discussed in [18]) and its deployment in WANs. This paper briefly discusses OpenFlow's role in WANs as a transformative approach to enhance flexibility, scalability, and innovation while integrating seamlessly with existing network architectures and technologies. Similarly, [17] offers a comprehensive examination of SDN, but the concept of SD-WAN is only briefly mentioned. The sole reference to WAN highlights the need for deploying multiple controllers to enhance reliability and reduce control plane latency. Furthermore, [19] delivers an in-depth survey on Transport SDN (T-SDN), addressing its application in transport networks. While the paper briefly introduces SD-WAN, it primarily characterizes it as an extension of SDN and Network Function Virtualization (NFV) principles applied to the edge of transport networks. Blenk et al. [20] conducted a survey that categorizes and explores Network Virtualization Hypervisors for SDN environments. Their work emphasizes the role of hypervisors in abstracting physical SDN networks into multiple logically isolated virtual networks, each with its own controller. They categorize hypervisors based on their architecture, centralized versus distributed, and their execution platforms, such as general-purpose compute platforms and network elements.

The evolution of optical transport SDN solutions has been a remarkable journey, with surveys tracing its path from the early stages. The inception of Software-Defined Optical Networks (SDON) marked a critical milestone, as researchers began integrating SDN concepts into single-domain optical networks, unifying control over optical and IP layers [21]–[24]. These early works paved the way for the subsequent stages of SDN development in the optical domain. To conclude this first group of SDN related surveys, Kankipati et al. [25] go in the direction of SDN for WAN, exploring its role in cloud connectivity and digital transformation, especially regarding security and network management. The authors outlined the key features and benefits of SD-WAN by emphasizing the flexibility, cost-effectiveness, and enhanced network performance it offers. Bannour et al. [26] provide a detailed and structured survey on distributed SDN control, presenting both a physical and logical taxonomy of SDN controller architectures. The paper emphasizes the scalability, reliability, and consistency challenges faced by centralized SDN deployments, and analyzes how various distributed controller platforms (e.g., ONOS, HyperFlow, Ravana, Kandoo) address these limitations. It offers comparative insights into controller placement, state synchronization, and fault-tolerance strategies. While

[2]For the sake of clarity, this Table I includes only the acronyms that appear more than once in the manuscript.

the discussion is comprehensive with respect to intra-domain and data center environments, the treatment of SD-WAN is indirect. The work mentions the need for distributed control in WAN-scale deployments but does not specifically delve into the unique characteristics, deployment models, or operational challenges of SD-WANs.

In Table II, we present a summary of these works, highlighting their contributions and limitations concerning SD-WAN.

Regarding the surveys that focus specifically on SD-WAN, we can notice that the literature is sparsely populated. While the existing works offer valuable contributions, they do not yet provide a fully comprehensive perspective on the emerging SD-WAN technology.

The survey in [27] provides a high-level overview of SD-WAN, identifying four primary research directions: traffic engineering, network optimization, systems, and service orchestration. While it introduces these key areas, the discussion remains broad and lacks a focused examination of practical deployment scenarios or detailed methodologies for implementing advanced TE techniques. Similarly, Michel et al. [28] present a concise review of SDN in WAN, emphasizing its benefits, such as improved load distribution, dynamic adaptation to network demands, and enhanced resilience to failures. However, this work primarily highlights advantages without delving into SD-WAN-specific challenges or solutions. The survey in [29] narrows its scope to SD-WAN architecture, providing an overview of its three-plane model (data, control, and application planes), but does not address practical considerations like network monitoring, real-world deployment strategies, or performance evaluation. Collectively, these surveys provide valuable insights but fall short in offering a comprehensive analysis of SD-WAN, particularly in areas such as detailed deployment scenarios, advanced traffic engineering methodologies, and network monitoring techniques.

The paper in [30] provides a thorough analysis of SD-WAN architecture, distinguishing between the control plane and data plane, with a particular focus on the CPE. While this survey evaluates performance aspects such as controller placement, TE, and load balancing, our work expands on these topics by categorizing deployment scenarios into three distinct groups: Managed Telco or Internet Service Provider (ISP) SD-WAN, Managed Cloud SD-WAN, and Unmanaged SD-WAN. This categorization provides a more nuanced understanding of SD-WAN deployment strategies. Additionally, while [30] addresses TE with a general perspective on performance metrics and load distribution, our survey delves deeper into traffic classification and policy-based routing algorithms, including heuristic-based and ML-based approaches. Furthermore, our work distinguishes itself by offering an in-depth analysis of network observation and monitoring techniques, including both active and passive methods, which are briefly mentioned in [30], and the inclusion of emerging research directions in in-band network telemetry for monitoring.

Table III provides a detailed comparison of how extensively various SD-WAN survey papers address each topic. To quantify the depth of investigation, we counted the number of words dedicated to each topic in each paper. Coverage is categorized as follows: $\star$ *limited* (fewer than 300 words), $\star\star$ *moderate* (300–600 words), and $\star\star\star$ *in-depth* (more than 600 words). This classification reflects the level of detail devoted to each topic.

However, we point out that significant opportunities for deeper exploration remain. While SDN has been extensively studied, a thorough examination of SD-WAN's innovative potential has yet to be fully realized. This survey aims to address this gap by offering a comprehensive analysis of this transformative technology, shedding light on its advancements and future possibilities.

A. Aim and organization of this survey

This survey is structured to provide a comprehensive understanding of SD-WAN, covering both foundational concepts and advanced topics. Section III introduces SD-WAN's fundamental principles and key components, establishing the groundwork for deeper exploration. Section IV examines the architectural requirements for an agile and efficient control plane, a cornerstone of SD-WAN functionality. Section V explores deployment scenarios and prototypes, illustrating SD-WAN's adaptability across various use cases. Key challenges and emerging research directions are discussed in Section VI. Within this context, Section VI-A focuses on traffic engineering techniques that enhance control plane efficiency, while Section VI-B reviews monitoring methodologies critical to ensuring seamless SD-WAN operations. Section VI-C addresses control plane resource management, including failure handling and controller placement strategies. The survey concludes with two final sections. Section VII synthesizes key insights from this study, summarizing lessons learned. Finally, Section VIII presents our perspective on the future of SD-WAN, identifying trends and opportunities for further research and development.

III. Introduction to Software-Defined Wide Area Network

In this section, we provide an introductory overview of SD-WAN by focusing on two fundamental aspects: its architecture and resource management components. These pillars are critical for understanding the core operation of SD-WAN, encompassing both the physical infrastructure and the deployment options that suit diverse network environments. Additionally, the software-driven intelligence that powers SD-WAN's agility and adaptability is explored through its resource management components. By clearly distinguishing between these pillars, we provide a comprehensive foundation that paves the way for more detailed discussions in later sections, where each aspect will be examined in depth.

SD-WAN architecture consists of multiple key components that enable efficient network management, optimized traffic routing, and enhanced application performance. As illustrated in Fig. 2, these components include the Customer Premises Equipment (CPE), the underlay access network, the overlay network, the controller, and the orchestration plane, all aligned with Metro Ethernet Forum (MEF) standards [31].

To illustrate these interactions, consider a large retail company with multiple branch offices. Each branch has CPE devices that manage traffic, security, and WAN optimization

TABLE II: Existing surveys on SDN

Related works	Topic	Key Contribution	Limitations w.r.t SD-WAN
[16]	SDN with OpenFlow	It provides an overview of SDN, highlighting implementations with OpenFlow and some hints of implementation in WAN.	It lacks focus on WAN, and the main focus is only related to OpenFlow.
[17]	SDN in different domains	It gives a comprehensive survey on SDN technologies across multiple network domains	The concept of SD-WAN is not introduced, and the reference to the use of SDN in WAN is limited.
[18]	SDN and OpenFlow	It provides an overview of SDN, highlighting implementations with OpenFlow	There are no references to the use of SDN in WAN.
[19]	SDN for transport Network	It gives a detailed look into Transport SDN	The references to SD-WAN and WAN are limited.
[20]	Network Virtualization Hypervisors for SDN	It categorizes and explores SDN hypervisors, providing detailed insights into their role in SDN infrastructure.	It focuses only on hypervisors, excluding SDN for WAN and SD-WAN.
[21]	SDN in optical transport networks	It investigates the integration of SDN in optical transport networks, focusing on OpenFlow-based solutions.	It focuses only on single-domain optical networks and does not extend to SD-WAN.
[22]	Evolution of Software Defined Optical Network	It discusses next-gen optical networks, focusing on programmable transceivers, automation, and integrating optical wavelengths with Optical Transport Network (OTN) and MPLS layers.	It does not consider SDN advancements in WAN connectivity.
[23]	Transport SDN with OpenFlow	It reviews how Transport SDN enhances resource allocation in optical transport networks.	There are no references to SDN for WAN or SD-WAN.
[24]	Optical transport SDN	It reviews SDN and its application to various networks, focusing on the separation of data and control planes for simpler resource management.	It lacks a cross-domain perspective and does not cover SD-WAN.
[26]	Distributed SDN control	Provides a comprehensive survey and taxonomy of distributed SDN controllers, analyzing their architectures (flat, hierarchical), logical models, and fault-tolerant capabilities.	The work offers a thorough analysis of SDN controller architectures; however, it does not explicitly address SD-WAN-specific challenges, deployment considerations, or integration within WAN environments.

TABLE III: Existing surveys on SD-WAN with: ⋆ limited, ⋆⋆ moderate, ⋆⋆⋆ in-depth analysis

SD-WAN Survey		[27]	[28]	[29]	[30]	This Survey
SD-WAN Architecture	CPE				⋆⋆	⋆⋆⋆
	Underlay			⋆	⋆	⋆⋆⋆
	Overlay			⋆	⋆	⋆⋆⋆
	Control plane		⋆	⋆	⋆	⋆⋆⋆
	Orchestration plane	⋆			⋆	⋆⋆⋆
Traffic Engineering	Traffic classification					⋆⋆⋆
	Performance metrics				⋆⋆	⋆⋆⋆
	Policy-based routing	⋆⋆	⋆	⋆	⋆⋆	⋆⋆⋆
Network Observation and Monitoring	Active	⋆				⋆⋆⋆
	Passive				⋆	⋆⋆⋆
	In-band network telemetry					⋆⋆⋆
Control Plane Resource Management	Failure management	⋆⋆	⋆		⋆	⋆⋆⋆
	Controller placement	⋆⋆	⋆⋆	⋆	⋆⋆⋆	⋆⋆⋆

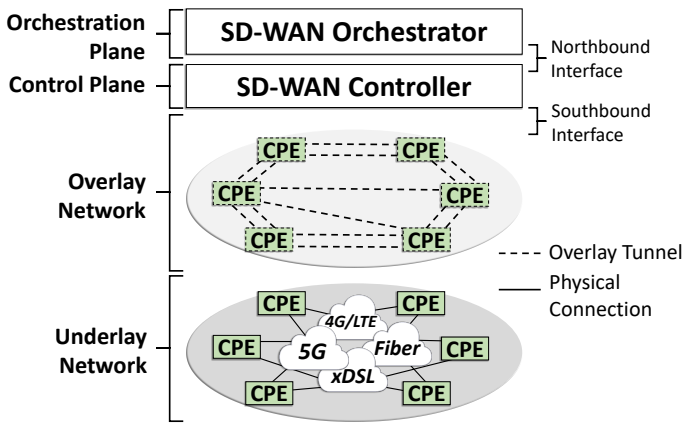

Fig. 2: SD-WAN high-level architecture, designed in alignment with MEF standards [31].

based on real-time network conditions. These CPEs connect to underlay access networks such as broadband Internet and 4G/LTE, forming the SD-WAN backbone. Over this, a virtualized overlay network abstracts the complexities of physical infrastructure, seamlessly interconnecting CPEs across locations for flexible and adaptive traffic control. The intelligence of SD-WAN lies in its control and orchestration planes. The controller, often termed the "brain" of the system, facilitates communication between CPEs and the orchestration plane, directing network traffic to optimize performance and reliability. Unlike traditional VPNs, SD-WAN's dynamic control logic ensures robust application performance even under network congestion or failures. Meanwhile, the orchestration plane hosts software and applications that define network policies via programmable interfaces.

Fig. 2 provides a high-level view of SD-WAN as a customizable framework that providers tailor to meet specific business

needs. SD-WAN providers fall into two broad categories:

- Managed SD-WAN Service Provider (MSP): Managed Service Providers control the core network, offering fully managed SD-WAN solutions covering design, deployment, monitoring, and maintenance. Major MSPs include telecom companies like AT&T and Verizon, as well as Cloud Service Providers (CSPs) such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud, which integrate SD-WAN into their cloud services.
- Unmanaged SD-WAN Service Provider (USP): Unmanaged Service Providers cater to organizations managing their own edge networks but relying on ISPs for Internet access. USPs provide flexible software and hardware solutions supporting self-managed SD-WAN deployments with features like dynamic path selection, application visibility, and cloud integration. Notable examples include Viptela (Cisco), VeloCloud (VMware), and Silver Peak.

Section V explores different SD-WAN deployment scenarios, analyzing how various service providers adapt solutions to meet specific organizational requirements.

SD-WAN resource management comprises integrated software-driven mechanisms that orchestrate traffic, optimize performance, and enhance security. These components enable real-time adaptability, centralized management, and comprehensive network visibility. The key elements of SD-WAN resource management include:

- Traffic Engineering: SD-WAN dynamically routes traffic across multiple underlay networks (e.g., broadband, MPLS, 4G/LTE) based on real-time network monitoring metrics such as latency, packet loss, and available bandwidth. It ensures high-priority applications receive necessary resources while optimizing lower-priority traffic, enhancing network efficiency and performance.
- Network Monitoring: Continuous tracking of key performance indicators (e.g., link status, throughput, latency) provides real-time visibility into network health. Advanced monitoring tools use AI and ML to detect anomalies, predict issues, and recommend optimizations to maintain seamless application performance.
- Control and Data Plane Resource Management: The control plane orchestrates data flow across the network, while the data plane executes routing decisions. Coordinated management ensures dynamic bandwidth allocation, optimal routing, and efficient application prioritization.
- Security and Threat Mitigation[3]: Integrated security features, such as encryption, firewalls, and Intrusion Detection and Prevention System (IDPS), protect data across public and private links. Centralized security management enforces policies consistently across sites, while SD-WAN can reroute traffic to avoid compromised paths, mitigating risks from threats like Distributed Denial of Service (DDoS) and man-in-the-middle attacks.

Each component is critical to maintaining SD-WAN's adaptability and responsiveness to evolving network demands, though challenges in their implementation must be addressed to maximize SD-WAN's potential. A primary challenge is traffic engineering, which must optimize traffic distribution while balancing congestion, latency, and application priorities, particularly when using best-effort broadband Internet access. Research focuses on developing advanced models for intelligent, real-time routing that considers QoS and Quality of Experience (QoE), ensuring reliable application performance.

Network monitoring is equally critical, as SD-WAN abstracts multiple underlay networks into a unified overlay. Traditional monitoring tools lack the granularity needed for visibility across transport layers. Advanced observation techniques leveraging AI and ML can improve real-time insights, predict network behavior, and support automated SD-WAN management, reducing operational overhead.

Security remains a major concern. SD-WAN's distributed architecture, reliance on public Internet access, and multi-cloud deployments introduce new attack vectors. Research in this area should focus on robust encryption, intrusion detection, and adaptive threat intelligence. Security frameworks must evolve to handle the complexity of dynamic, multi-site environments, ensuring data integrity and network resilience.

These areas are deeply interconnected. Effective traffic engineering relies on accurate monitoring to enable real-time adjustments, while robust security measures must maintain network performance and visibility. Enhancing one function often strengthens the others, for instance, optimized traffic management reduces congestion and minimizes associated vulnerabilities, while improved monitoring enhances security policies, enabling faster threat detection and mitigation. This synergy underscores the importance of a holistic approach to SD-WAN design, where performance, security, and adaptability work in unison to create a resilient and intelligent network infrastructure. Advancements in these domains will lead to more reliable, efficient, and secure SD-WAN deployments. Optimized traffic engineering ensures consistent application performance across diverse underlays, intelligent monitoring enables proactive issue resolution, and advanced security safeguards sensitive data. Together, these improvements will drive broader SD-WAN adoption across industries. Section VI explores these resource management functions in detail, examining their role in SD-WAN's evolution and highlighting research directions to overcome their challenges.

IV. Network Architecture

SD-WAN architecture can be categorized based on the degree of control over the data plane, distinguishing between Core and Edge control models. Fig. 3 shows these models by offering different approaches to managing and optimizing network traffic, each with its own set of advantages.

The **Core control model** involves the integration of SD-WAN capabilities directly into the existing network infrastructure of the SD-WAN provider. This approach incorporates SD-WAN functionalities into the provider's existing routers or switches, as well as the CPEs deployed at the customer's branches. As elaborated in the following sections, the Core model is commonly adopted by traditional telecommunications companies and cloud providers. It offers particular advantages

[3]We emphasize that this survey does not address SD-WAN security, as it merits a dedicated study.

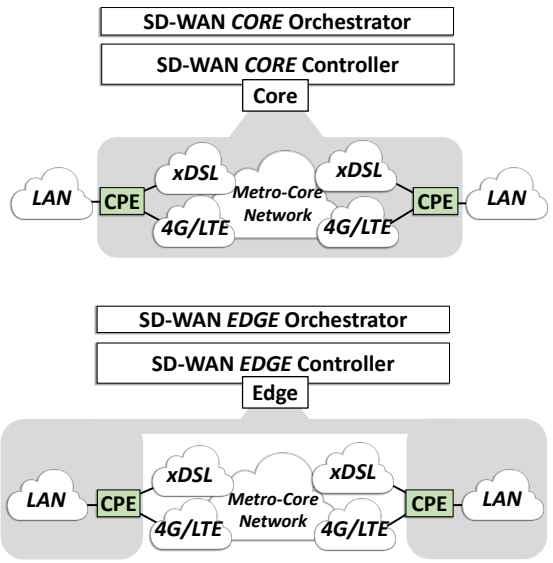

Fig. 3: SD-WAN Core and Edge control models.

to organizations with well-established network infrastructures looking to embrace SD-WAN while making use of their existing networking investments.

In contrast, the **Edge control model** focuses on deploying SD-WAN exclusively at the network's edge, specifically within CPE devices located in branch offices, headquarters, and DC sites of enterprises. Consequently, we will refer to this approach as SD-WAN Edge, as it operates independently of the provider's network infrastructure by enabling enterprises to manage and customize their edge networks. This model often leads to cost savings by reducing the need for expensive MPLS connections and allowing for more efficient use of cheaper Internet connections. The deployment of these two models introduces distinct challenges for SD-WAN providers that must be carefully addressed to ensure optimal performance and seamless integration. For the Core control model, providers face the complexity of integrating SD-WAN functionality into their existing network infrastructure, which requires careful planning to avoid disruptions and ensure compatibility with legacy systems. This can also involve managing the balance between leveraging existing hardware and rolling out new features that demand higher processing power or more advanced networking capabilities. On the other hand, the Edge control model poses challenges around managing distributed deployments, particularly in terms of ensuring consistent performance across multiple sites, maintaining central control while allowing localized flexibility, and addressing concerns about security, redundancy, and latency. SD-WAN providers must also tackle issues related to controller placement and the architecture's scalability, ensuring that the solution can accommodate growing network demands across diverse geographical locations. Details of these deployment challenges will be further explored in the subsequent sections.

A. Customer Premises Equipment (CPE)

The CPE serves as the edge device that enables remote sites, such as branch offices or DCs, to seamlessly connect to other branches and, by extension, to the broader company or DC network. Traditional CPEs are typically hardware-based devices, often proprietary "black boxes" with fixed configurations and network interfaces provided by ISPs to deliver communication services. However, as businesses require more advanced functionalities, such as load balancing, encryption, firewalls, antivirus protection, and other network services, these traditional CPEs become increasingly complex and expensive to upgrade and manage.

To address the limitations of traditional hardware-based CPEs, a shift has occurred towards virtualizing many of the functions that used to be hardware-dependent. Instead of housing all network services within a physical CPE device, Network Function Virtualization (NFV) enables these services to run as software, decoupling them from the hardware. This transformation allows for standard off-the-shelf CPE hardware to be used, which reduces cost and increases flexibility. As a result, software-based Virtual Network Function (VNF)s can replace hardware components, such as routers, firewalls, load balancers, and switches, and run on standard operating systems within a DC or other virtual environments [32].

With NFV, the CPE's functionalities are no longer confined to hardware and can be deployed in different forms depending on the business needs. Two key variants have emerged in SD-WAN deployments:

- virtual CPE (vCPE): In this model, the CPE's functionalities are fully virtualized and hosted in the cloud or a DC. By running these functions remotely, businesses can centralize the management of their network services, reducing the need for complex hardware at each site. This approach allows for greater scalability and rapid deployment of new services, as functions can be updated or modified without physical changes at the remote sites. However, since traffic between remote sites may need to traverse the central location where the virtualized functions are hosted, this can introduce a forwarding triangle, potentially increasing latency and reducing efficiency for certain traffic flows.
- universal CPE (uCPE): The uCPE model, on the other hand, allows businesses to deploy off-the-shelf, generic hardware at the remote site, which can run virtualized network functions locally. This gives businesses the flexibility to customize services on-site, such as firewalls or routing, while still benefiting from the adaptability and lower cost of software-based solutions. The uCPE can host multiple VNFs simultaneously, providing a more versatile and dynamic edge device that adapts to new services without requiring new hardware.

With both vCPE and uCPE, businesses have the option to decide whether they want to run network functions in the cloud (vCPE), on-site (uCPE), or a combination of both [33], depending on their specific needs. This flexibility makes SD-WAN deployments more cost-effective, scalable, and adaptable to evolving network requirements [34] [35] [36]. The main functions provided by an SD-WAN CPE include:

- Traffic routing and optimization: an SD-WAN CPE can intelligently select the best path for network traffic based

on real-time monitoring of network conditions, such as latency, packet loss, and available bandwidth. It can also distribute network traffic across multiple WAN links to maximize utilization and minimize congestion while prioritizing critical applications. Moreover, it can automatically switch traffic to alternate paths in case of link failures to ensure continuous connectivity.

- Security, application visibility, and control: SD-WAN CPE devices include firewall capabilities to protect the network from external threats and support encryption to secure data in transit, particularly when using public Internet links. Some CPEs offer Intrusion Detection and Prevention (IDP) features to detect and mitigate security threats. These devices can inspect packet payloads using Deep Packet Inspection (DPI) to identify and classify applications, then direct traffic based on application priorities and policies.
- WAN optimization and link aggregation: SD-WAN CPEs can enhance network performance by compressing data to reduce bandwidth consumption, caching frequently accessed data locally, and optimizing Transmission Control Protocol (TCP) traffic over high-latency links. For example, in satellite networks, where latency fluctuates, SD-WAN can optimize TCP by using a proxy at the network edge. This proxy splits the TCP connection, hiding remote delays and improving TCP performance by enabling faster congestion window growth. These devices can aggregate multiple WAN links (e.g., MPLS, broadband, 4G/LTE) into a single logical connection to increase bandwidth and reliability.
- Centralized management and zero-touch provisioning: SD-WAN CPEs can be centrally managed from an SD-WAN control plane framework, allowing for easy configuration, monitoring, and policy enforcement across multiple sites. They also support zero-touch provisioning, simplifying the initial setup and configuration process.

While the CPEs are essential for the operation and management of SD-WAN, their effectiveness largely depends on the quality and capabilities of the underlay access network to which they connect. The underlay access network provides the essential physical and transport foundation that enables the overlay networks of SD-WAN to function optimally.

B. Underlay access network

In SD-WAN architectures, the underlay access network connects CPEs to ISP gateways, providing the foundation for SD-WAN's flexible overlay. Enterprises deploy various technologies based on service level requirements, including broadband (xDSL, fiber-optic), cellular (4G/LTE, 5G), and satellite communications.

1) Broadband technologies: xDSL, and fiber-optic connections are widely used in SD-WAN for cost-effectiveness and flexibility, offering an alternative to MPLS links. A retail chain, for instance, may leverage broadband to connect stores to a central DC, while fiber-optic ensures high bandwidth for applications like point-of-sale systems and inventory management. Broadband supports bandwidth-intensive applications such as video conferencing and cloud services. A technology firm may use broadband to facilitate remote work, ensuring seamless access to cloud-hosted tools. However, the lack of QoS guarantees can impact real-time applications. Financial institutions managing low-latency trading platforms may experience congestion-related performance issues, necessitating SD-WAN traffic prioritization, as discussed in Section VI-A. Security risks arise when using public broadband networks, requiring encryption and firewall measures, as detailed in Section IV-C. For example, hospitals relying on broadband for telemedicine must comply with the Health Insurance Portability and Accountability Act (HIPAA) regulations [37] by implementing secure communication protocols. Broadband is also valuable in disaster recovery. A logistics firm may use broadband as a backup connection to maintain supply chain operations during network failures.

While cost-effective [38], [39], broadband can suffer from congestion, making SD-WAN's intelligent routing crucial for stability. Despite its advantages, broadband's reliability may not suffice for industries requiring minimal service interruptions. A gaming company hosting live eSports events, for instance, may find variable latency unacceptable. In such cases, broadband is best complemented with MPLS or cellular connectivity for improved resilience. Broadband's affordability and performance make it ideal for SD-WAN applications, provided its limitations are mitigated through real-time analytics, traffic shaping, and failover mechanisms.

2) Cellular technologies: 4G/LTE and 5G play a crucial role in SD-WAN, particularly for mobility, rapid deployment, and failover. Logistics firms use 4G/LTE for fleet communication, enabling real-time GPS tracking and route optimization [40], [41]. Similarly, emergency response teams rely on cellular networks to establish command centers in areas lacking wired infrastructure [42], ensuring real-time coordination. Cellular networks also enable temporary healthcare clinics in remote areas to access patient records and conduct remote consultations.

However, challenges include data caps and high costs for data-intensive applications. Retailers using cellular as a primary connection for point-of-sale systems and security cameras may face increased expenses if they exceed data quotas. Additionally, congestion in urban areas can impact performance. Latency and bandwidth fluctuations also affect financial applications. Traders using 5G to access trading platforms remotely may experience delays due to network congestion. To address such issues, SD-WAN incorporates dynamic bandwidth allocation. A construction company using 4G/LTE for remote job sites can prioritize critical project data over non-essential traffic. Seamless handover between cellular and other networks is essential for continuous connectivity. SD-WAN enables moving trains providing onboard Wi-Fi via 5G to switch towers without interruptions. Similarly, retail stores using cellular backup during wired outages benefit from automatic failover to primary connections once restored. While 4G/LTE and 5G offer deployment speed and mobility, their constraints necessitate careful traffic management and integration with other underlay technologies. Various industries demonstrate their potential in SD-WAN when combined with

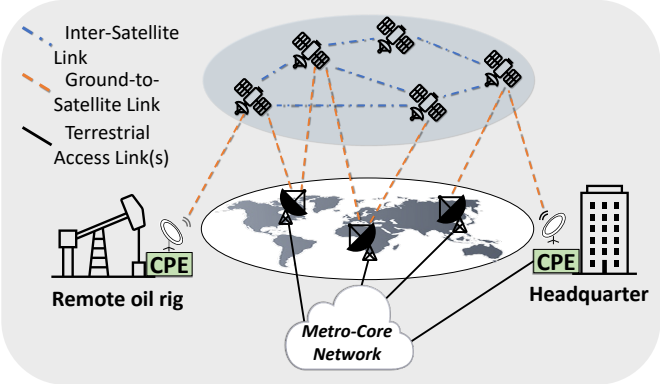

Fig. 4: Example of satellite-based SD-WAN linking an oil rig with company headquarters.

intelligent traffic routing and prioritization.

3) Satellite connections: The integration of satellite connectivity into SD-WAN architectures is transforming industries operating in remote or dynamic environments. Recent advancements in Low Earth Orbit (LEO) satellite constellations have significantly reduced latency and improved bandwidth compared to traditional Geostationary Orbit (GEO) satellites, making them a viable option for global connectivity. A study in [43] comparing real and simulated satellite measurements highlights LEO's lower round-trip time, enhancing SD-WAN performance in long-distance operations.

A key application is the maritime industry, where ships rely on satellite connectivity for navigation, communications, and real-time data exchange. SD-WAN enables dynamic switching between LEO satellite links and terrestrial networks as ships approach ports, optimizing costs and performance. Versa's SD-WAN deployment for a global cruise company [44] exemplifies this, automating transitions between satellite and terrestrial links, eliminating manual intervention, and optimizing bandwidth allocation. With SD-WAN, multiple satellite beams with varying latencies and bandwidths can be intelligently managed through traffic shaping and load balancing, ensuring a seamless Internet experience for crew and passengers.

Similarly, the aviation sector benefits from SD-WAN-enabled satellite networks, supporting in-flight Internet services and operational communications. SD-WAN's ability to transition between satellite and terrestrial links ensures uninterrupted service on global flight routes.

In remote industries, SD-WAN enhances satellite-based Internet access. Oil and gas companies connect offshore platforms to onshore control centers [45], ensuring real-time monitoring of drilling operations (see Figure 4). Disaster response teams leverage SD-WAN with satellite links to establish communication networks in regions where terrestrial infrastructure is damaged.

Satellite integration extends SD-WAN coverage to areas beyond terrestrial networks, benefiting logistics, mining, and agriculture [46]. Satellite links also serve as resilient failover connections, ensuring business continuity. Financial institutions in remote areas, for instance, use satellite networks as backup links for uninterrupted access to critical applications. Despite its advantages, satellite connectivity presents challenges. Latency and bandwidth fluctuations caused by weather conditions [47] affect real-time applications like video conferencing and Voice over IP (VoIP). SD-WAN mitigates these issues by prioritizing latency-sensitive traffic through adaptive routing algorithms. However, effectiveness varies based on implementation. The limited bandwidth of satellite connections necessitates efficient traffic management. For example, an e-commerce company using satellite links for warehouse operations must prioritize inventory updates over non-essential data transfers.

Cost is another significant limitation. LEO satellite services remain costly, with Gbps-month pricing being 1000 times higher than that of submarine cables [48]. SD-WAN solutions must implement intelligent traffic policies to balance performance and costs. A research station in Antarctica, for instance, could reserve satellite links for critical data uploads while scheduling non-urgent updates during off-peak hours. Advancements in LEO satellite constellations like Starlink and OneWeb address some challenges, offering lower latency and higher throughput [49] [50]. SD-WAN solutions designed for these networks enhance performance and reliability, enabling hybrid setups that combine LEO satellites with terrestrial underlays for optimal connectivity. One novel approach is proposed in [51] which focuses on proactively mitigating LEO link latency variations. The system steers mission-critical traffic to terrestrial underlays before anticipated latency spikes occur. This predictive capability uses a lightweight satellite constellation simulator embedded within the SD-WAN edge device. The simulator forecasts potential path changes and their performance impact.

Emerging multi-orbit SD-WAN architectures leverage LEO, Medium Earth Orbit (MEO), and GEO satellites simultaneously to improve connection resilience [52] [53]. GEO satellites provide broad coverage with high latency, MEO offers a balance between latency and coverage, while LEO ensures low-latency, high-bandwidth performance. The integration of satellite networks into SD-WAN unlocks new opportunities for industries in remote environments, offering global reach, enhanced reliability, and operational flexibility. Overcoming latency, bandwidth, and cost constraints requires advanced SD-WAN features such as traffic prioritization, compression, and adaptive routing. As satellite technology evolves, its role in SD-WAN will continue to expand, driving innovation across diverse sectors. While the underlay network provides foundational connectivity, the overlay network defines SD-WAN's intelligence, enabling advanced routing, traffic optimization, and security features.

C. Overlay network

The fundamental distinction between a traditional WAN connectivity service and one delivered via SD-WAN hinges on the concept of overlay networking. The traditional WAN services are provided exclusively by ISPs. They encompass various types, including Layer 1 connectivity [54], Carrier Ethernet [55], and IP/MPLS services [56], and are implemented using a diverse array of network technologies.

Therefore, this underlying connectivity is not directly controlled by the subscriber, such as an enterprise or business

customer, but it is often controlled and maintained by the WAN service provider (e.g., ISPs).

An overlay network is defined as a virtual network constructed on top of one or more underlying network infrastructures, enabling seamless connectivity between CPEs. As we will show in Section V, the precise configuration of the overlay network is contingent upon the type of SD-WAN service being deployed. For instance, an enterprise's overlay network might depend on a VPN managed either by an SD-WAN provider with a Core control model (e.g., MPLS-VPN) or directly by the customer with an Edge control model. In the Edge control model, the VPN is constructed atop the public Internet, leveraging one or more Internet access networks active at the customer premises.

As such, the overlay networking technologies can be divided into two categories: provider-provisioned and customer-provisioned. The provider-provisioned category is based on various technologies that are realized and maintained by SD-WAN providers and are, therefore, proprietary. Meanwhile, customer-provisioned solutions include overlay networking techniques and protocols that are often open-source and built/maintained only by the organization. In the following paragraphs, we provide an overview of overlay networking technologies in this category, such as Virtual eXtensible Local Area Network (VXLAN), Generic Routing Encapsulation (GRE), and IP security (IPsec).

1) Virtual eXtensible LAN (VXLAN): VXLAN is an encapsulation protocol that provides end-to-end connectivity, such as for DC networks, using tunneling to stretch L2 connections over an underlying L3 network [57]. It can be considered as an evolution of the well-known Layer 2 Tunneling Protocol (L2TP) defined in [58] [59], which was used to extend the Point-to-Point Protocol (PPP) [60] model by allowing L2 and PPP endpoints devices to reside on different locations interconnected by an L3 network.

Nowadays, VXLAN is the most commonly used protocol to create overlay networks in DC networks due to its scalability benefits. Over the past decade, DCs have rapidly increased server virtualization, leading to a significant need for agility and flexibility in terms of allocation of network and computing resources. VXLAN represents a technology that allows the segmentation of DC L2 networks, but it also solves the scaling limitation of Virtual LANs (VLANs) by increasing scalability to up to 16 million logical networks (as opposed to 4094 VLANs).

The VXLAN tunneling protocol encapsulates L2 Ethernet frames into L4 User Datagram Protocol (UDP) segments (using 4789 as default port), as in L2TP, by enabling the creation of virtualized L2 subnets that span over physical L3 networks. Each L2 subnet is uniquely identified through a 24-bit segment ID, called Virtual Network Identifier (VNI) that segments traffic of Virtual Machine (VM). Differently from L2TP, the encapsulation and decapsulation of packets is performed by the VXLAN Tunnel End Point (VTEP) that is located within the hypervisor on the server that hosts the VMs. Thus, the VNI and VXLAN-related tunnel outer header encapsulation are known only to the VTEP and not to the VMs. Moreover, VTEPs could also be on a physical switch or physical server and could be implemented in software or hardware.

In a nutshell, VXLAN offers the following benefits:

- As mentioned, VXLAN supports a maximum of 16M VXLAN segments with 24-bit VNIs so that a DC can accommodate numerous tenants.
- It reduces the number of MAC addresses that network devices need to learn and enhances network performance because only uCPEs devices at the edge of the VXLAN network need to identify VM MAC addresses.
- Extends L2 networks using MAC-in-UDP encapsulation and decouples physical and virtual networks. Tenants are able to plan their own virtual networks without being limited by the physical network IP addresses or broadcast domains. This greatly simplifies network management.

VXLAN integration into SD-WAN presents a compelling solution for organizations requiring scalable, flexible, and efficient network segmentation across distributed infrastructures. By supporting up to 16 million VNIs, VXLAN enables the logical separation of traffic, which is crucial for large-scale deployments like those of multinational corporations managing extensive branch networks or cloud service providers hosting multi-tenant environments. The protocol's decoupling of physical and virtual networks allows seamless connectivity over geographically dispersed locations, making it particularly advantageous in scenarios such as global retail operations or DC interconnects, where consistent architecture and efficient resource allocation are critical. Additionally, VXLAN reduces the burden on core network devices by confining MAC address learning to edge devices, enhancing overall performance.

However, the benefits of VXLAN come with challenges. The encapsulation process adds overhead to packet sizes, which can lead to bandwidth inefficiencies in constrained environments like satellite-based SD-WAN setups, where every bit of capacity matters. Deploying VXLAN also demands meticulous configuration of VTEPs and VNIs, increasing operational complexity for organizations with limited technical expertise or those relying on legacy hardware that may not support VXLAN. Furthermore, the protocol's reliance on a stable underlying network exposes it to disruptions caused by packet loss or high latency, which can critically impact performance for applications requiring low latency and high reliability, such as real-time analytics or telemedicine. Security is another concern, as VXLAN does not inherently provide encryption, necessitating additional measures like IPSec to ensure data protection, further increasing deployment complexity. Despite these hurdles, advanced SD-WAN capabilities such as dynamic routing, intelligent traffic prioritization, and integrated security mechanisms enable organizations to mitigate these challenges. When effectively deployed, VXLAN allows enterprises to harness its scalability and flexibility, transforming their network architectures into robust, future-proof systems capable of meeting evolving demands.

2) Generic Routing Encapsulation (GRE) and IP security (IPsec): GRE is a tunneling protocol initially developed by Cisco Systems, with its specifications defined in RFC 1701 [61]. It serves to encapsulate a diverse array of network layer protocols, ensuring a private path for packet delivery over an

IP network. A fundamental aspect of GRE is its ability to encapsulate the original IP packet, typically generated by the Local Area Network (LAN) at the edge, within an outer IP packet. Consequently, GRE tunnel endpoints transmit original IP packets through GRE tunnels, facilitating their routing to edge networks. Upon reaching the endpoint, GRE tunnel encapsulation is stripped away, allowing the IP packet to proceed toward its final destination.

The encapsulation procedure operates as follows:

- When the CPE receives a packet from a host of the edge LAN to be tunneled, it sends the packet to the tunnel interface.
- The tunnel interface encapsulates the packet in a GRE packet and adds an outer IP header.
- The IP packet is forwarded on the basis of the destination address in the outer IP header.

The decapsulation procedure operates as follows:

- When the CPE receives the IP packet from the tunnel interface, the outer IP header and GRE header are removed.
- The original packet is routed towards the destination edge network based on the inner IP header.

One of the key advantages of GRE is its protocol agnosticism. It can encapsulate a wide range of network layer protocols, including IPv4, IPv6, and non-IP protocols. This flexibility makes it suitable for SD-WAN, which may involve the encapsulation of diverse types of traffic over public Internet connections. GRE can be used to establish VPNs by encapsulating traffic and ensuring it remains separate from other network traffic. Moreover, GRE is widely supported by many networking devices and operating systems, making it a versatile choice for creating tunnels between different SD-WAN devices and across diverse network environments. GRE tunnels can connect remote sites seamlessly and support dynamic routing protocols. In an SD-WAN context, this allows for intelligent traffic routing and load balancing based on real-time network conditions, a crucial feature for optimizing connectivity.

While GRE itself does not provide encryption, it can be combined with other encryption protocols, such as IPsec, to enhance security. This combination offers data confidentiality and integrity for traffic traversing the tunnel, which is important in SD-WAN deployments, especially when data is transmitted over untrusted networks like the public Internet. IPsec is a secure network protocol suite that authenticates and encrypts data packets to provide secure encrypted VPNs between two hosts, or networks, over an IP network. IPsec includes protocols for establishing mutual authentication between end-to-end hosts in WANs, as such it can protect data flows through the negotiation of cryptographic keys to use during the VPN session [62]. It supports network-level peer authentication, data origin authentication, data integrity, and data confidentiality (encryption). IPsec can operate in two modes: transport and tunnel. In transport mode, only the payload of the IP packet is encrypted or authenticated, therefore, the routing over the IP network works normally. While in tunnel mode, the entire IP packet is encrypted and authenticated. It is then encapsulated into a new IP packet with a new IP header. Tunnel mode is used to create VPNs for network-to-network communications (e.g., enterprise networks), host-to-network communications (e.g., remote user access), and host-to-host communications (e.g., private chat) [63].

GRE and IPsec play complementary roles in SD-WAN deployments, each offering distinct advantages and addressing critical requirements for modern enterprise networks. GRE, as a protocol-agnostic tunneling technology, provides the flexibility to encapsulate various network layer protocols, including IPv4, IPv6, and even non-IP traffic. This versatility is particularly valuable in SD-WAN environments where diverse types of traffic may need to traverse public or private networks. For instance, GRE tunnels can be used to seamlessly connect remote branch offices to a central DC, enabling dynamic routing protocols to operate efficiently across geographically dispersed locations. The simplicity and broad compatibility of GRE make it an excellent choice for organizations that require interoperability across different networking devices and environments.

However, GRE itself lacks native encryption, which limits its ability to secure data transmitted over untrusted networks like the public Internet. This is where IPsec becomes essential. By combining GRE with IPsec, SD-WAN deployments can achieve both the flexibility of GRE and the robust security features of IPsec. IPsec ensures data confidentiality, integrity, and authentication through its encryption and key exchange mechanisms, making it indispensable for securing sensitive traffic. For example, a financial institution deploying SD-WAN can use GRE to encapsulate dynamic routing traffic while relying on IPsec to encrypt and authenticate data transmissions between branch offices, safeguarding critical financial transactions and customer information.

The integration of GRE and IPsec provides significant advantages, but it also presents challenges. GRE encapsulation adds minimal overhead, which is beneficial for optimizing bandwidth usage. However, when combined with IPsec, the additional encryption headers can increase packet size, potentially impacting performance in bandwidth-constrained environments. For instance, in a satellite-based SD-WAN deployment, the combined overhead of GRE and IPsec may strain limited link capacity, necessitating careful traffic optimization and prioritization.

Another challenge lies in the complexity of managing combined GRE and IPsec configurations, particularly in large-scale SD-WAN deployments with numerous sites. Setting up and maintaining tunnels, managing cryptographic keys, and ensuring consistent policy enforcement across diverse environments require advanced tools. For example, an enterprise with hundreds of branch locations might face operational hurdles in coordinating GRE and IPsec configurations across its entire network.

Despite these challenges, the combination of GRE and IPsec remains a cornerstone of SD-WAN architectures. GRE enables seamless and protocol-agnostic connectivity, while IPsec provides the robust security needed to protect data in transit. Together, they empower organizations to create scalable, flexible, and secure network infrastructures, supporting applications

that range from secure remote access for employees to resilient inter-site communications for global enterprises.

Although IPsec and GRE are two of the most popular technologies for creating overlay networks in SD-WAN, innovative techniques are being proposed in multi-site interconnection scenarios on a global scale. New SD-WAN architectures based on dynamic overlays exploit the integration of distributed points of presence (PoPs) within cloud edge networks and are proposed as alternatives [64]. Unlike IPsec VPN, these systems allow more granular control of traffic with centralized management of such overlays, allowing traffic to be dynamically diverted through less congested paths, improving both reliability and overall network performance.

D. Control plane

As illustrated in Fig. 2 and Fig. 3, the control plane bridges the orchestration plane and the data plane made by overlay and underlay network layers. The software-based control plane enables programmatic access to network resources and forwarding policies. It specifically establishes and dynamically manages overlay tunnels across data plane devices. These include technologies like VXLAN, GRE, and IPsec (detailed in Section IV-C). This control makes network administration agile and flexible [65]. It is considered the "brain" of the entire architecture that controls the communication between applications in the orchestration plane (such as business and enterprise services) and data plane devices, such as CPEs and WAN nodes. We can identify two types of SD-WAN controllers based on the type of control model: Edge and Core. The first model aims to control and monitor the customer network, including CPEs and end-to-end overlay tunnels. The core model is responsible for controlling and monitoring the provider network. It integrates SD-WAN capabilities directly into the SD-WAN provider's existing network infrastructure (routers, switches, etc.).

Both types of controllers include two main functionalities [66], namely network control, i.e., the set of functions that contribute to configuring and maintaining the state of the network, and network monitoring, i.e., the set of functions aimed at monitoring the health of the network. In the following, we delve into the specific functions and design of both controller types.

1) Edge controller: As mentioned in the previous paragraph, the Edge model focuses on controlling the customer network, i.e., handling the CPEs and the overlay networks. The network control functions comprise techniques for CPEs configuration and maintenance. It ensures proper provisioning, software updates, and policy compliance. Moreover, it establishes and monitors virtual network overlays (e.g., GRE [61], IPsec [62]), optimizing traffic routing and resource allocation. Additionally, it enforces traffic policies, prioritizing applications, selecting optimal routes, and balancing loads dynamically. Simultaneously, the SD-WAN edge controller provides real-time monitoring of CPE status, performance, and connectivity, enabling proactive troubleshooting. It tracks overlay network metrics (latency, packet loss, bandwidth) to optimize application delivery and dynamically adjust traffic routing. Additionally, it maintains an updated network map, ensuring efficient data flow and path selection.

2) Core controller: The Core model focuses on controlling the provider network, i.e., handling the WAN nodes and the mapping between the underlay and overlay networks. The network control functions comprise techniques for provisioning and maintaining WAN nodes, handling setup, updates, and configurations in line with network policies. It optimizes both underlay and overlay networks to ensure efficient, secure application delivery. Additionally, it implements policy-based routing for traffic prioritization, dynamic route selection, and load balancing. Simultaneously, the controller collects real-time data on WAN node performance and connectivity, enabling proactive troubleshooting. It monitors key metrics (latency, packet loss, bandwidth) within the underlay network to optimize traffic routing and ensure high application quality. Additionally, it maintains an updated network topology map for efficient routing and management.

Several open-source controllers can be employed in SD-WAN, each crucial in managing network functions. Ryu [67], Open Network Operating System (ONOS) [68], OpenDaylight [69], Floodlight [70] and the new Teraflow [71] are the most employed. Ryu controller (Python-based) provides developers with defined Application Program Interface (API) software components, facilitating the seamless creation of novel network management and control applications. It supports diverse protocols for managing network devices, including but not limited to OpenFlow, Netconf, and OF-config. Such protocols also permit the configuration of overlay tunnels and associated forwarding rules. ONOS controller (Java-based) offers scalability and high availability, making it suitable for large-scale network deployments. It simplifies creating and deploying dynamic network services with user-friendly interfaces and ensures simplified management, configuration, and deployment of software, hardware, and services through a modular, distributed control. OpenDaylight (Java-based), a collaborative project under the Linux Foundation, is a modular SDN controller platform. With a robust ecosystem and diverse set of plugins, OpenDaylight facilitates the development of SDN applications, catering to different network architectures and requirements. Floodlight (Java-based), an Apache-licensed SDN controller, is known for its stability and performance. It is widely used in research and industry applications, providing a solid SDN development and experimentation foundation. Teraflow SDN controller [71] (Python-based) is an innovative, cloud-native SDN controller designed to advance network automation and management. It features a microservices-based architecture, enabling seamless integration with NFV and MEC (Multi-access Edge Computing) frameworks. Teraflow offers advanced capabilities in flow management and network equipment integration, incorporating ML-based security measures and policy-driven forensic evidence for multi-tenancy. Its modular design ensures scalability and resilience, making it suitable for complex, high-performance network environments.

These controllers leverage a suite of standardized southbound protocols, such as OpenFlow [72], Netconf/YANG [73], and Restconf [74]. OpenFlow is a widely adopted protocol that facilitates communication between the Controller and

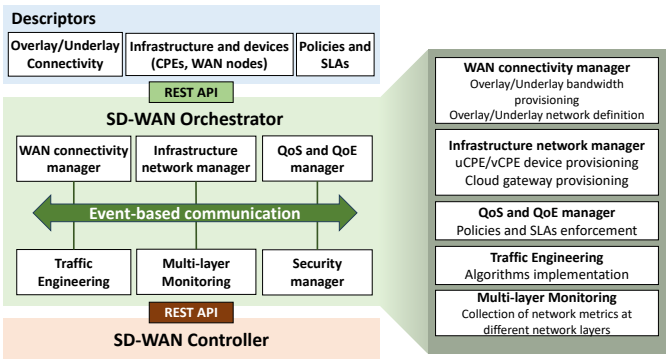

Fig. 5: SD-WAN orchestration plane.

CPE/WAN devices. Through OpenFlow, the controller can configure CPEs, manage flow tables, and make dynamic decisions about how network traffic is forwarded. Netconf (Network Configuration Protocol) and YANG (Yet Another Next Generation) data modeling language provide a standardized framework for configuring and managing network devices. The Controller utilizes Netconf/YANG to interact with CPE/WAN devices, defining configurations, setting parameters, and ensuring consistency across the network. This approach streamlines provisioning and facilitates seamless firmware and software updates. Restconf is a RESTful web service protocol that extends Netconf capabilities to web-based applications. It provides a lightweight, HTTP-based interface for configuring and managing network devices. The Controller employs Restconf to establish a user-friendly and web-based interaction with CPE/WAN devices.

E. Orchestration plane

The orchestration plane hosts the software applications that define business or enterprise network policies. These applications leverage the programmability of the controller's Network-Based Interface (NBI) and Virtual Infrastructure Managers (VIMs): through the NBI, they gather infrastructure data, such as the state of overlay networks, while through VIMs, they manage VNFs at both the customer level (e.g., CPEs) and the provider level (e.g., cloud environments). By modifying CPE policy rules based on company needs, these applications govern overall SD-WAN behavior. They rely on an abstracted network view from the controller to implement control logic, which the controller then translates into low-level network instructions.

To frame the discussion, the following paragraphs review key open-source orchestrators for SDN/NFV. While no equivalent open-source platform currently exists for SD-WAN, these solutions provide foundational components that future SD-WAN orchestrators are likely to build upon, as further discussed later in this section.

In the context of open-source solutions, several projects have emerged to address network orchestration challenges. One such initiative is the Open Source MANO (OSM) project, hosted by ETSI, which aims to provide an open-source orchestrator closely aligned with the ETSI NFV reference architecture [75]. OSM serves as the NFV Orchestrator (NFVO) and facilitates the deployment and management of VNFs through a set of descriptors, albeit requiring some level of expertise for effective use. Another notable project, the Open Network Automation Platform (ONAP), is under the Linux Foundation's umbrella and strives to enable end-to-end service agility across various domains using a unified platform for NFV and SDN orchestration, fully adhering to ETSI directives [76]. ONAP incorporates modules for big data and AI, facilitating policy optimization and automated network service deployment and management. However, it demands substantial installation resources and is more complex to deploy than OSM. Additionally, there's Cloudify, an open-source cloud orchestration software platform that has been extended to encompass NFV-related use cases [77]. While Cloudify implements most of the ETSI NFV MANO functional blocks, it falls short of full compliance due to its incompatibility with legacy systems. These orchestration platforms primarily operate on descriptors, which define network services in terms of VNF components, instantiation parameters, and VNF Forwarding Graph (VNF-FG) specifications. The responsibility for managing cross-site connectivity within these platforms lies with the WAN Infrastructure Manager (WIM).

However, it is essential to note that WIM remains partially defined, with no currently available open-source implementation. Many commercial SD-WAN solutions provide different features and adapt to different use cases. In this section, we focus on investigating which are the main ingredients of an SD-WAN orchestration system as shown in Fig. 5.

The SD-WAN orchestrator is conceived as a modular system, breaking down the network orchestration and control platform into discrete building blocks, each serving a specific role within the system [78]. The orchestration modules, which assume the role of the SD-WAN Orchestrator, operate at a higher level of abstraction, avoiding direct interactions with the network equipment to be managed [79] [80]. In this model, the intents and configurations of WAN domains received through descriptors are translated into specific operations. These operations can either remain internal or be directed toward the controller, which maintains a direct connection to the infrastructure. This connection allows them to apply the necessary configurations to the network infrastructure, aligning it with the declared requirements in the intent.

The WAN connectivity manager module plays an important role in maintaining seamless WAN connectivity by provisioning overlay and underlay bandwidth to optimize resource allocation across the network infrastructure. A framework introduced in [81] enables WAN virtualization for dynamic end-to-end bandwidth provisioning, focusing exclusively on WAN connectivity without addressing interactions with cloud domains or NFVOs for comprehensive connectivity management. Another approach leverages a WAN Infrastructure Connectivity Manager (WICM) to integrate VNFs with WAN connectivity services [82], simplifying resource management while providing an abstract view to NFVOs.

The infrastructure network manager oversees the provisioning and integration of uCPE/vCPE devices and cloud gateways within the network. An architecture design proposed in [83] enables orchestration for these devices in multi-site environments, outlining key components, interfaces, and APIs

essential for multi-site 5G infrastructure. Another approach in [84] introduces a Data Plane Broker (DPB) supporting multi-domain connectivity, which integrates with orchestration systems through a WIM plugin, enabling rapid WAN setup and efficient resource allocation.

The QoS and QoE manager, together with the TE and multi-layer monitoring modules, enforce policies and SLAs to guarantee the QoE of network services. Thanks to the TE algorithms aimed at optimizing network traffic routing and resource allocation, and software tools for real-time visibility into network performance, we ensure that network performance aligns with predefined service level agreements and QoS standards. Multi-layer monitoring provides comprehensive visibility across infrastructure, network, and service layers, enabling correlation of performance metrics and early detection of SLA violations, thereby empowering the QoS/QoE manager and TE algorithms to make informed decisions for dynamic traffic management and policy enforcement. OSM can be used as an enabler for the deployment of Network Services (NSs) across multiple interconnected domains [85], handling VNF interconnections within domains by meeting a pre-defined QoS policy [86].

An open-source SD-WAN orchestrator has been presented in [78]. It offers a flexible, resilient, and cloud-native orchestration solution tailored for enterprise and academic networks. This solution is built entirely on open-source tools, leveraging ONOS as the underlying controller [68]. The orchestrator is designed as a modular system, decomposing the network orchestration and control platform into smaller, functional building blocks, as illustrated in Fig. 5. Each module adheres to a consistent implementation paradigm, employing a microservices architecture to enhance scalability and maintainability.

From an operational perspective, the modules expose REST endpoints to enable interaction with external components, such as descriptor provisioning. Simultaneously, they utilize REST clients to integrate with the controller's functionality. Within this REST-based framework, the orchestrator manages the core operations of each module. These operations involve receiving input from REST endpoints, performing necessary actions (potentially involving event-driven communication with other modules), and sending directives to the controller to implement the desired network configurations. This modular and service-oriented design ensures flexibility and adaptability for evolving network requirements.

Each SD-WAN provider offers various architectural network models to cater to the diverse networking needs of organizations. These architectural approaches define how SD-WAN components are deployed and interconnected within a network. In the following section, we delve into the different SD-WAN deployment scenarios introduced by different service providers.

V. Deployment Scenarios

The SD-WAN landscape is characterized by its dynamic nature, with continuous innovation, changing market dynamics, and evolving customer requirements. As a result, the classification of SD-WAN service providers, together with their network architecture and deployment scenarios, can shift over time based on various factors. These factors include advancements in technology, emerging business models, and evolving customer preferences/needs [87]. Nowadays, it is important to note that providers can be primarily categorized into two distinct types:

- **MSP**: These providers offer fully managed SD-WAN solutions, providing end-to-end services, including design, deployment, monitoring, and management of the SD-WAN infrastructure. The key feature of these providers is that they have control of their core network, and as a result, they can implement the core control model into their commercial solutions. In this sense, they can offer an SD-WAN service to their customers that is very well optimized by meeting the customers' requests. Examples of managed SD-WAN service providers include:
  - Traditional Telco and ISP: Traditional telecommunications companies and ISPs often offer SD-WAN services as an extension of their existing network services. For instance, AT&T, Verizon Communications, Deutsche Telekom, Nippon Telegraph & Tel (NTT), etc., provide managed SD-WAN solutions with a focus on reliable connectivity and integration with their broader network offerings.
  - Cloud Service Provider (CSP): Major cloud providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) offer SD-WAN solutions that are optimized for connecting to their cloud services. These providers often focus on enhancing connectivity to cloud resources and applications.
- **USP**: These providers specialize exclusively in offering solutions for the management of edge networks. These solutions are particularly tailored for companies that lack a core network connecting their remote branches but have access to the Internet via ISPs. USPs offer software and hardware-based SD-WAN solutions that empower organizations to establish and oversee their own SD-WAN deployments. By facilitating the creation of end-to-end virtual tunnels, these solutions enable the interconnection of various branches. USPs are recognized for their innovative and adaptable approach in delivering SD-WAN solutions that align with the unique needs of different organizations. They prioritize features such as advanced application visibility, dynamic path selection, and seamless cloud integration. This empowers organizations to optimize network performance for a range of workloads and applications, allowing them to embrace SD-WAN while leveraging their existing network infrastructure. Prominent examples of Unmanaged SD-WAN vendors include Viptela (acquired by Cisco), VeloCloud (acquired by VMWare), Silver Peak, and others.

The choice of the SD-WAN service provider depends on factors such as the organization's size, budget, existing network infrastructure, geographical locations, desired level of control, and specific business needs. Each type of provider

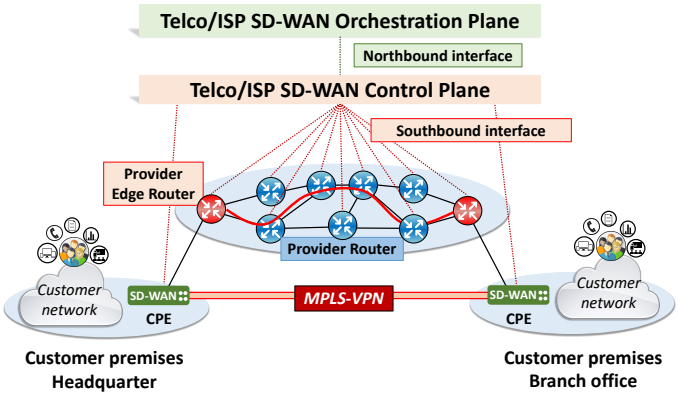

Fig. 6: Telco/ISP SD-WAN architecture model.

brings its own advantages and considerations to the table, making it important for businesses to carefully evaluate their options before making a decision.

A. Managed telco/ISP SD-WAN

Telecommunications companies have the ability to integrate SD-WAN capabilities directly into their existing network infrastructure. This approach involves incorporating SD-WAN functionalities into existing routers or switches to both manage the operator's network and deliver SD-WAN services to customers.

Fig. 6 shows an example of an SD-WAN architecture implemented by a Telco. The SD-WAN Control and Orchestration planes manage Telco's network infrastructure and facilitate remote configurations of routers and switches dedicated to serving businesses and enterprises. Specifically, we can state that the Telco/ISPs can implement the SD-WAN Core control model to deliver WAN services to their customers. They can deploy the SD-WAN functionalities in the customer's CPEs to coordinate and manage the connectivity services.

Unlike traditional WANs, Telcos may leverage SDN technology to automate configuration processes for network devices through a dedicated management network often owned by the provider (Northbound and Southbound interfaces). Telco-supplied CPE acts as black boxes, connecting company offices to the provider's edge router. Within the SD-WAN CPEs, the Telco configures networking functions such as establishing overlay tunnels (e.g., MPLS), proprietary monitoring, and traffic engineering mechanisms [88].

Telco-driven SD-WAN implementations offer a range of valuable advantages. Leveraging their extensive networking expertise and deep experience, Telcos deliver reliable and optimized SD-WAN services to businesses. This expertise is complemented by their global network reach, enabling the provision of SD-WAN solutions across diverse geographical locations, even in remote areas. Telcos ensure carrier-grade performance, high availability, and low latency for crucial applications. Additionally, managed SD-WAN services from Telcos often come with SLAs that guarantee uptime, performance, and swift response times. The scalability of Telco-driven SD-WAN solutions allows them to adapt to changing business needs, facilitating growth and expansion. In the next paragraph, we survey the related work on architecture prototypes for managed Telco/ISP SD-WAN services.

1) Overview of architecture prototypes for managed Telco/ISP SD-WAN services: The design of SD-WAN architectures in Telco/ISP environments is important for balancing priority and best-effort traffic while ensuring SLA compliance.

A prototype tested in a real network [89] uses Juniper Network's OpenLab [90] and Northstar controller [90] to evaluate a managed Telco SD-WAN solution. The architecture leverages REST APIs [91] for monitoring and configuration, while a proprietary Southbound interface manages device-level operations. The controller orchestrates IP/MPLS flows, automates traffic engineering, and provides a GUI for network monitoring. Testing involved eight routers across multiple U.S. states with IXIA routers generating realistic traffic [4].

Other solutions explore OpenFlow-based SD-WAN architectures using simulated [92] and emulated [93] environments. [93] integrates the FloodLight controller [70] with Open vSwitch WAN nodes, while [94] introduces a Protocol-Oblivious Forwarding (POF)-based SD-WAN for scalable WANs. [92] presents SafeGuard, a Python-based congestion-aware failure recovery mechanism implemented on the Ryu SDN controller [67] and evaluated on Google B4 [95] and AT&T [88] topologies using Mininet [96].

Further studies examine Telco/ISP SD-WAN performance [97]–[100]. [97] tests load distribution on simulated and real topologies from Rocketfuel [101] and Abilene[5]. [98] applies Abilene for traffic engineering, while [99] evaluates a Fast ReRoute QoS protection approach on a 16-router, 24-link network. [100] tests real ISP topologies, including BellCanada from the Internet Topology Zoo [102] and a custom SD-WAN testbed. Integrating Segment Routing over IPv6 (SRv6) enhances SD-WAN by enabling network slicing, differentiated SLAs, and cloud-integrated virtual network elements. SRv6 allows efficient traffic engineering through end-to-end tunnels [103], simplifying path allocation via real-time performance metrics. It reduces state complexity and supports Service Function Chaining (SFC) for advanced network services. However, SRv6 introduces a header overhead associated with using long IPv6 addresses, which increases packet size. This can be mitigated using Generalized Segment Identifier (G-SID) compression [103] or alternative algorithms like HCMD [104], which reduces bandwidth waste from segment lists (more details can be found in Section VI-A4). For SRv6-based SD-WANs, network slicing is crucial. [105] implements a solution where SD-WAN edge routers classify traffic into end-to-end slices, functioning as L3VPNs. However, SRv6 adoption depends on IPv6 networks supporting Segment Routing Header (SRH) processing, which may be restricted by transport network policies. Performance monitoring solutions like Simple Two-way Active Measurement Protocol (STAMP) (detailed in Section VI-B) can help measure SRv6 path metrics. Finally,

[4]IXIA routers consist of high-performance multiport traffic generators/analyzers used for testing multilayer Packet OverSynchronous Optical NETworking (SONET), Gigabit, and 10/100 Mbps Ethernet switches, routers, and networks. Source: https://www.fiberopticsonline.com/doc/ixia-1600-0001 (last access: June 4th, 2025).

[5]Abilene topology and traffic information is available at SDNLib: https://sndlib.put.poznan.pl/home.action (last access: June 4th, 2025)

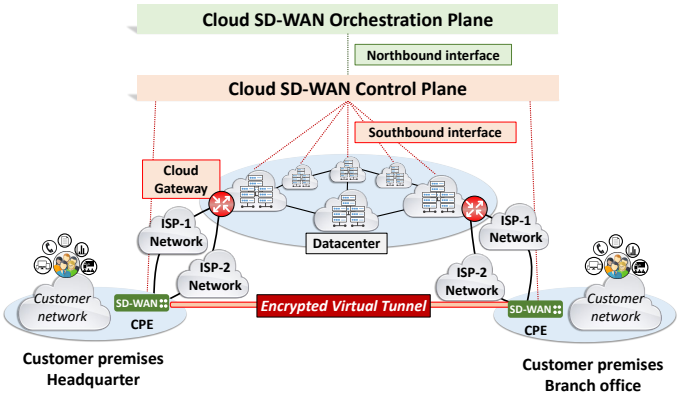

Fig. 7: Cloud SD-WAN architecture model.

EveryWAN [106] is an open-source prototype implementing SRv6 in SD-WAN. It uses Linux-based edge routers managed by a centralized controller, supporting network slicing, performance monitoring, and real-time traffic management.

B. Managed cloud SD-WAN

CSPs offer a distinct variant of SD-WAN service referred to as "Managed Cloud SD-WAN." This approach involves the seamless integration of the SD-WAN solution with cloud services and infrastructure, delivering uninterrupted connectivity between branch offices and cloud-based applications for customers [66]. As illustrated in Fig. 7, CSPs gain benefits from a hybrid architecture resembling the Core control model, similar to the Telco SD-WAN scenario. Therefore, it is important to note a key distinction: in this context, CSPs may not exercise control over the access network. Instead, customers procure Internet access through one or more ISPs to connect their CPEs to the respective cloud gateway. This distinction underscores the unique arrangement in which CSPs operate within the SD-WAN framework. As such, the CSP can capitalize on the SD-WAN features to exercise control over its own network of DCs, affording full governance. Leveraging their extensive global network infrastructure, CSPs optimize the performance and delivery of applications hosted across various DCs and cloud platforms. This type of SD-WAN holds significant advantages for businesses that heavily rely on cloud services, ensuring consistent and optimized connectivity for their widely dispersed workforce.

Fig. 7 depicts an illustrative Cloud SD-WAN architecture. This solution encompasses an on-site SD-WAN CPE that establishes connectivity with a cloud gateway. For instance, it facilitates real-time traffic shaping and the utilization of multi-circuit load balancing/failover capabilities. This is made possible by allowing the CPEs to harness one or more Internet accesses through diverse ISPs. Additionally, this setup contributes to the heightened performance and reliability of cloud applications. The cloud gateway is directly connected to the cloud services, resulting in an overarching enhancement of the overall performance of cloud applications. Furthermore, suppose the organization possesses alternative Internet access. In that case, the SD-WAN CPE can seamlessly reroute network traffic to the company's alternate Internet access, thereby preventing any disruption to ongoing sessions [107].

Two notable implementations of Cloud SD-WANs, designed by over-the-top CSPs Google and Microsoft, are worth highlighting: Google's B4 [108] and Microsoft's SWAN [109]. These initiatives showcase how leading companies are leveraging SD-WAN to enhance their Core network architectures and deliver optimized services to their customers.

Google's B4 [108] is a well-known SD-WAN solution developed by Google, designed to improve the efficiency and performance of Google's internal Core WAN that connects its DCs and provides connectivity for its services and applications. The architecture of B4 SD-WAN centers on several core principles, each contributing to its operational functions. By prioritizing dynamic traffic engineering, B4 maximizes efficiency across Google's global DCs, dynamically selecting optimal data paths in response to real-time conditions, thus minimizing latency and resource wastage. This emphasis on load balancing fosters superior performance and redundancy as sophisticated algorithms distribute traffic across multiple paths. Path Diversity emerges as a strategy to circumvent bottlenecks and congestion by simultaneously leveraging multiple network paths, ensuring consistent performance even during traffic spikes. The system's centralized control, orchestrated by the B4's central controller, assumes a central role in optimizing network traffic. Informed by various factors, including link quality, capacity, and congestion, this controller makes intelligent routing decisions, exemplifying Google's commitment to efficiency and network optimization.

Microsoft's SWAN [109] is a Cloud SD-WAN solution developed by Microsoft to improve the performance, reliability, and efficiency of its global Core network infrastructure. SWAN aims to provide seamless connectivity and optimized data transmission across Microsoft's DCs, offices, and services, ensuring a better user experience for both internal users and external customers. The architecture of SWAN is composed of multiple layers and components, exploiting a combination of physical network devices, virtual appliances, and centralized controllers for effective network traffic management and control. SWAN places a strong emphasis on intelligent routing, application-aware traffic steering, and the efficient utilization of network paths, all geared towards enhancing the overall performance and reliability of the network. One notable feature of SWAN is its dynamic traffic rerouting capability, which operates based on real-time network conditions and application needs. SWAN can autonomously detect issues such as network congestion, link failures, or latency problems and make routing decisions to ensure optimal network performance.

In the next paragraph, we survey the related work on architecture prototypes for managed Cloud SD-WAN services.

1) Overview of architecture prototypes for managed cloud SD-WAN services: While B4 and SWAN demonstrate how leading CSPs leverage their vast resources to enhance their Core architectures, other studies explore Managed Cloud SD-WANs in a broader range of scenarios, often with experimental or open-source approaches. For example, a typical operational scenario is presented in [110], where three branch offices are interconnected via two WANs. Using Mininet [96] and Open vSwitch to emulate OpenFlow-enabled switches and hosts in customizable topologies, the testbed relies on the

FloodLight controller to deliver OpenFlow functionalities and a development environment for network services. Another experimental SD-WAN testbed scenario that uses the FloodLight controller is implemented in [111] to assess the network's performance while providing VoIP services. The architecture integrates virtualized components across two physical servers through VMware, with Mininet [96] providing the data plane and OpenDayLight [69] serving as a redundant controller for failover. Further extending this line of work, [9] utilizes Open^3++ [85], an extension of Open Source MANO [75], OpenStack, and OpenDaylight to support multi-domain NS deployment. An Orchestration proxy manages VNF connections between these domains, interacting with the WIMs responsible for the network domains that interconnect the compute domains. An SDN controller represents each WIM and utilizes OpenDaylight and the Virtual Tenant Manager (VTN) plugin for SD-WAN implementation. Another open-source SD-WAN solution is presented in [112], in which the authors employed two Mininet [96] virtual machines running on Ubuntu 18.04 LTS. They configured three instances of Open vSwitch within the first VM to ensure robust network availability. Similarly, the second Mininet VM was equipped with three OVS switches to facilitate LAN communication. The SD-WAN controller, utilizing OpenFlow 1.4, dynamically updates the flow tables of each Open vSwitch. Inter-network connectivity was established through a full-mesh configuration of GRE tunnels [113], complemented by an additional full-mesh arrangement of VXLAN [114] overlay tunnels. To enhance control plane reliability, they implemented a cluster comprising two Docker containerized instances of ONOS [68], ensuring increased availability. Each Mininet instance emulated two hosts to evaluate inter-cloud performance, assessing various scenarios. These scenarios encompassed (1) evaluating seamless connectivity and (2) comparing the effective data rate and Round-Trip Time (RTT) between GRE and VXLAN tunnels.

C. Unmanaged SD-WAN

The Unmanaged SD-WAN architecture model consists of a virtual network layer that operates independently on top of the existing underlay physical infrastructure. SD-WAN CPEs are deployed at the network's edges, typically at branch offices or remote sites of organizations. These devices establish secure tunnels over the underlying network, enabling centralized control and management of traffic flows. This model offers enhanced visibility, traffic optimization, and application performance control [66]. It allows organizations to prioritize specific applications, apply security policies, and dynamically route traffic based on real-time conditions. In this context, the various technologies to realize WANs have become a commodity for customers who require a connectivity service to their remote locations. The customers have the freedom to add or remove different kinds of WAN connectivity services (Internet broadband, MPLS, Carrier Ethernet) based on their needs.

Fig. 8 depicts an illustrative Unmanaged SD-WAN architecture that incorporates an on-site SD-WAN CPE responsible for

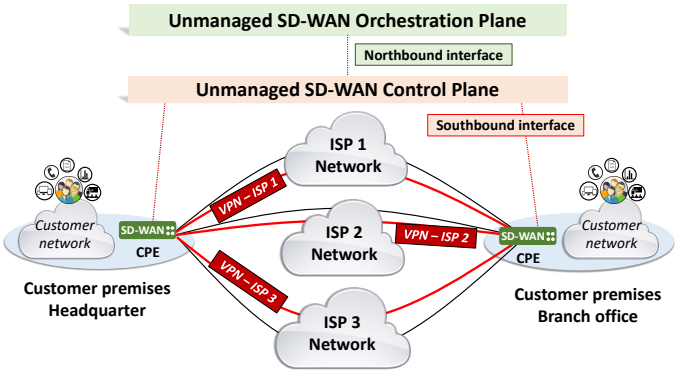

Fig. 8: Unmanaged SD-WAN architecture model.

establishing VPN tunnels with other branch CPEs. WAN connections are facilitated through overlay VPNs configured over diverse underlay Internet/broadband access networks delivered by ISPs. The control plane assumes a central role, guided by the SD-WAN controller, which translates instructions from the orchestration plane into actions performed by the CPEs in the data plane. With awareness of the CPEs' functional models, the controller adeptly transforms high-level traffic engineering decisions from the orchestration layer into appropriate CPE instructions. Communication between CPEs and the controller transpires through the South-Bound Interface (SBI), employing dedicated SBI protocols. The orchestration plane is the network's intelligence center positioned atop the control plane. Hosted applications within this plane act as software programs facilitating policy communication and resource requests with the controller via NBI.

In the next paragraph, we survey the related work on architecture prototypes for unmanaged SD-WAN services.

1) Overview of architecture prototypes for unmanaged SD-WAN services: Advancements in mobile networks have driven SD-WAN integration, particularly in unmanaged deployments. In [115], an SD-WAN prototype with one CPE and five WAN links utilizes OpenAirInterface [116] to emulate real-world scenarios, testing video streaming over 4G/LTE connections. Similarly, [117] presents a virtual WAN with VyOS routers and Routing Information Protocol (RIP) [118], using Open vSwitch [119] for overlay configuration and D-ITG [120] for traffic generation and monitoring. VPN tunnels secured with IPsec ensure traffic integrity, while OpenDaylight [69] serves as the SD-WAN controller. [121] introduces a VXLAN-based dynamic tunnel approach to reduce network congestion and improve user access, tested using Mininet. EveryWAN [106] is an open-source SD-WAN implementation supporting multi-WAN overlays and automated network management. It includes the EveryEdgeOS SD-WAN controller for tenant registration, authentication, and service provisioning, with testing conducted using the EveryWAN Validation and Performance Evaluation (EVPE) platform. Another SD-WAN prototype in [122] integrates HTTP/2 prioritization into the network layer using Mininet, Open vSwitch, and the Ryu controller [123], [124]. The SD-WAN controller dynamically routes high-priority traffic over the lowest-delay path. In [125], QoS and performance were analyzed in an Indonesian company's

transition to SD-WAN, comparing MPLS, Ethernet over IP, and IPsec tunnels between Jakarta and Surabaya. Internet Control Message Protocol (ICMP) echo tests measured latency, jitter, and packet loss under varying traffic conditions.

[126] evaluates two SD-WAN testbeds using a Transport-Layer Monitoring (TPM) system: one in a research lab and another at an Italian municipality. The lab testbed employs ONOS, OpenFlow switches, Raspberry Pi-based CPEs, and GRE tunnels, while the municipal testbed simulates city hall and branch office connectivity. The experiments measure recovery time and service availability. An ONOS-based SD-WAN testbed in [127] uses Mininet and OpenFlow switches to simulate transit Autonomous System (AS) networks, leveraging Border Gateway Protocol (BGP) Quagga routers. The topology is generated using the Dorogovtsev-Mendes Algorithm [128], and Iperf is used for traffic testing. Towalink [129] is an open-source SD-WAN solution using WireGuard VPN [130] for secure site interconnection. Deployed across eight sites, it demonstrates resilience to outages and network failures. Similarly focused on open-source implementation and performance, [131] presents an SD-WAN CPE software leveraging VPP/DPDK for high-speed packet processing and WireGuard for secure tunneling. This solution was evaluated within a real metropolitan testbed. [132] presents an SD-WAN prototype utilizing Open vSwitch, VXLAN, and generic WAN links, focusing on Multi-Path TCP (MPTCP)-based WAN aggregation and congestion recovery. [133] models an SD-WAN with a headquarters and multiple remote sites, deploying six tunnels for bidirectional traffic. The aggregate traffic pattern generated on each tunnel adheres to a diurnal cycle and encompasses video flows. Metrics, including maximum link utilization, end-to-end delay, and end-to-end packet loss, are measured and compared against traditional aggregation algorithms like ECMP.

Table IV summarizes the research works mentioned in the previous subsections by highlighting their main implementation features. It is worth noting that the architectural models typical of Managed and Unmanaged SD-WAN providers are not the only representations of this technology. Given the inherent flexibility of SD-WAN, the concept of hybrid SD-WAN deployments emerges as a way to optimize network connectivity by integrating components from various SD-WAN solutions. These scenarios harness the advantages of different deployment models to accommodate a wide range of business needs and network landscapes. Whether integrating public cloud services, maintaining existing MPLS connections, or accommodating a combination of on-premises and remote locations, hybrid SD-WAN strategies provide adaptable solutions that balance performance, cost-effectiveness, and flexibility.

VI. Challenges and Research Directions

As SD-WAN continues to gain traction across various industries, its adoption presents several challenges that highlight the urgency and significance of further research in this area. Although SD-WAN delivers substantial benefits in terms of flexibility, cost efficiency, and performance optimization, several technical and operational issues must be addressed to unlock its full potential. These challenges encompass a range of areas, such as traffic engineering, network monitoring, control and data plane resource management, and security.

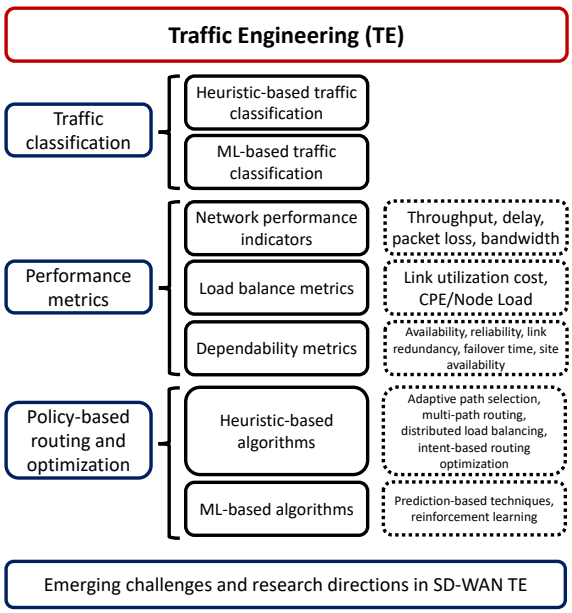

Fig. 9: Taxonomy of traffic engineering.

To ensure the long-term success and scalability of SD-WAN, targeted research is required to address these challenges, improve overall performance, and enhance security. Moreover, the interdependencies between these research areas must be carefully considered to develop holistic solutions that improve SD-WAN deployments.

A. Traffic Engineering (TE)

SD-WAN has revolutionized WAN management, offering agility, enhanced application performance, and cost efficiency. However, efficient traffic management and routing remain key challenges.

Researchers are developing intelligent algorithms to dynamically optimize traffic flows based on application requirements, network conditions, and cost constraints. TE algorithms enhance real-time adaptability by integrating key features such as traffic classification, network monitoring, and policy-based routing. The following sections explore these features and their role in optimizing SD-WAN performance, as summarized in Fig. 9.

1) Traffic classification: It refers to the ability to categorize network traffic based on attributes such as application type, source, destination, and user identity. In SD-WAN, this capability enables the TE algorithm to differentiate between various traffic types and apply specific policies accordingly.

Several methods can be employed for traffic classification [142] with some commonly used in SD-WAN environments, including Heuristic and ML-based classification. The choice of a particular classification method depends on the specific needs of the SD-WAN deployment, the required level of granularity, and the available computational resources.

a) Heuristic-based traffic classification: SD-WAN uses heuristic-based traffic classification to categorize network

TABLE IV: SD-WAN Prototypes

SD-WAN service type	Testbed	CPE	Overlay Tunnel Technology	Controller
Managed Telco/ISP	Real: [89], [100], [103], [134], [135]	Custom: [89], [100], [103], [134], [135]	-	NorthStar: [89]
	Simulated: [99], [98], [100], [97], [92]	Openflow switch: [92] Custom: [99], [98], [100], [97],	-	Ryu: [92]
	Emulated: [93], [94]	Openflow switch: [93], [94]	-	FloodLight: [93]
Managed Cloud	Real: [108], [109]	Openflow switch: [108] Custom: [109]	IP in IP: [108]	FloodLight: [109] Onix: [108]
	Simulated: -	-	-	-
	Emulated: [110], [111], [136], [9], [112]	Openflow switch: [110], [136], [112] Custom: [111], [9]	GRE: [136], [112] IPsec: [9] VXLAN: [112]	FloodLight: [110] OpenDayLight: [111], [136], [9] ONOS: [112]
Unmanaged	Real: [123], [126], [137], [131], [138], [139]	Openflow switch: [123], [126], [137], [138] Custom: [131]	IP over MPTCP: [123] GRE: [126], [137] IP over UDP: [129] [131]	Ryu: [123] ONOS: [126], [137], [138]
	Simulated: [133], [140]	Openflow switch: [133] Custom: [140]	-	-
	Emulated: [115], [117], [132], [78], [122], [141], [124], [127], [106], [121]	Openflow switch: [117], [132], [78], [122], [141], [124], [127], [106], [121] Custom: [115]	GRE over IPsec: [117], [78] VXLAN: [132], [106], [121]	OpenDayLight: [117] ONOS: [127], [78] Ryu: [122], [124]

flows based on IP addresses, port numbers, protocols, and other attributes. Implemented as a software application within CPEs, this method is managed by the orchestrator, which communicates with network controllers via APIs for intelligent traffic management. When a new data flow is detected, the forwarding logic captures the source and destination IP addresses [93]. If the destination is unknown, the frame is broadcast; otherwise, routing is optimized using a shortest path algorithm for resource efficiency. This process supports load balancing and QoS enforcement, ensuring high-priority traffic, such as VoIP and video conferencing, receives low-latency, high-bandwidth paths, while lower-priority flows, like file transfers, use cost-effective alternatives.

SD-WAN also classifies traffic based on network protocols (TCP, UDP, ICMP, HTTP, FTP, SIP), enabling WAN link aggregation for resilience and minimizing performance degradation [132]. Additionally, protocol-aware classification, such as HTTP/2 prioritization [122], enhances service delivery by dynamically adjusting traffic handling. These classification techniques enable intelligent load balancing, security enforcement, and QoS optimization, making them essential for managing complex and distributed SD-WAN environments.

b) ML-Based traffic classification: ML models are widely used for network traffic classification, leveraging packet headers, payloads, and behavior patterns to make data-driven decisions [143]. Supervised and unsupervised ML approaches improve classification accuracy and efficiency compared to traditional methods. Commonly used algorithms include K-nearest neighbor, Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression, and Decision Trees, with applications extending to wireless networks [144]. In SD-WAN, the Decision Tree algorithm enables CPEs to categorize packets in real time. This supervised model recursively splits data into subsets based on feature values (e.g., source and destination IP addresses, port numbers, total bytes sent and received, and protocol type) [141]. The algorithm makes binary decisions at each node (e.g., "Is the port number greater than 1024?" or "Is the protocol TCP?") to classify traffic such as file transfers and ping requests. This structured approach enhances packet analysis and routing decisions in SD-WAN environments. The Decision Tree algorithm offers interpretability and computational efficiency, making it well-suited for real-time SD-WAN applications. However, it is prone to overfitting, especially with deep trees, which can reduce generalization accuracy in dynamic traffic environments. Additionally, training and maintaining deep trees on large datasets can be computationally expensive. To mitigate these challenges, ensemble methods such as Random Forest and Gradient Boosted Trees improve generalization by combining multiple weak learners. Neural networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can model complex traffic patterns and temporal dependencies. RNNs are particularly useful for learning network states from historical data, enabling controllers to predict traffic conditions and optimize path selection [145]. Additionally, unsupervised methods like K-means clustering and semi-supervised approaches help handle unlabeled data, which is common in real-world networks.

2) Performance metrics: The TE algorithms implemented in SD-WAN must continuously monitor network conditions such as link quality, latency, packet loss, congestion events, and others. Real-time data collection is essential to make informed routing decisions. Moreover, incorporating QoS metrics related to different applications, such as minimum bandwidth and end-to-end delay constraints, is crucial to ensure that the selected path meets the performance requirements of different applications and services. In this section, we focus on enumerating the network metrics most commonly employed by state-of-the-art TE algorithms in SD-WAN. A more in-depth exploration of monitoring methodologies and algorithmic investigations will be the subject of Section VI-B.

a) Network performance indicators: Network performance indicators, often referred to as Network Key Performance Indicators (Net-KPIs), are essential metrics used to evaluate and monitor the efficiency, reliability, and overall health of an SD-WAN. These indicators include network throughput, delay, packet loss, bandwidth, and jitter.

Throughput quantifies data transmission volume over a network, measured in bps, Mbps, or Gbps. In SD-WAN, it assesses the network's ability to transfer data across branch offices, DCs, and cloud services. A key factor is the time required to restore throughput after a disruption or policy change [132], from the moment a link becomes undesirable to when the TCP connection fully recovers. Throughput measure-

ments help in reducing inter-cloud transit costs for ISPs [136], minimizing maximum link utilization [133], and mitigating WAN failures to prevent service disruptions [124], [132]. Additionally, they contribute to lowering bandwidth costs [146], further enhancing network efficiency, performance, and reliability.

Delay refers to the time required for data packets to travel from source to destination, typically measured in milliseconds (ms). In SD-WAN, it directly impacts application responsiveness and user experience, particularly for real-time services like video conferencing, VoIP, and online gaming. Jitter, or packet delay variation, affects real-time communication quality, leading to disruptions such as choppy audio or video. A key performance factor in SD-WAN is the time needed to restore delay to an acceptable range after an outage or policy change [132]. This period starts when a disruption occurs and ends when all lost packets are successfully retransmitted over a stable overlay tunnel. Since optimizing latency is crucial for maintaining application performance and user satisfaction, delay measurements play a critical role in SD-WAN. They support fault-tolerant reactive routing based on delay thresholds [89], help minimize end-to-end network delay to improve service application performance [97], [115], [147]–[149], and contribute to reducing service unavailability [117].

Packet Loss refers to the unintended and incomplete transmission of data packets between network endpoints within the SD-WAN infrastructure. It occurs when some packets sent from a source do not successfully reach their intended destination due to various factors, such as network congestion, errors, or disruptions along the underlay network. Packet loss measurements are used in various optimizations, such as improving bandwidth usage and minimizing packet loss through the implementation of Fast ReRoute schemes [99]. Additionally, these measurements help reduce end-to-end delay while maximizing throughput [150] and enhance the reliability of high-priority services [151]. A variation of packet loss used in SD-WAN is to calculate the Traffic Excess Factor (TEF) metric [152], representing the difference (in %) between the traffic of the single user and the available bandwidth normalized.

Bandwidth estimation refers to the available capacity on a network link, whether physical or virtual. In SD-WAN, TE algorithms take bandwidth availability into account to prevent congestion and ensure that applications receive the necessary resources. Bandwidth estimation is used to minimize service unavailability, optimize bandwidth usage, and reduce costs [93], [98], [108], [110], [134], [146].

b) Load balancing metrics: Besides the well-established network metrics, the load balancing metrics are essential components of SD-WAN management and optimization. Since these metrics are not as standardized as the previous ones, and their definition is open, we provide some examples for calculating them in SD-WAN scenario:

Link utilization cost metric measures the link utilization cost on each network link. These metrics help SD-WAN controllers assess the use of a specific link while also considering the associated cost. An approach for the cost of a link could be combining a fixed cost associated with using a specific network interface for the link and a cost that varies based on the amount of bandwidth utilized [100]. The fixed cost represents the price the ISPs charge for keeping the link operational. The second component reflects the fee ISPs pay for the bandwidth consumed when data is transmitted over the link. Together, these costs are in a linear relationship where the total cost increases as more traffic flows through the link, in addition to the baseline fixed cost for access. This metric is also used to reduce route length and link utilization for service applications [10] and optimize the implementation of routing policies [153].

CPE/Node Load metric evaluates the load on individual CPEs (in an SD-WAN Edge model) or intermediate nodes (in an SD-WAN Core model). In the former deployments, this metric encompasses the load on the edge devices, such as CPU utilization, memory usage, and the number of active connections. These factors are critical in assessing whether a CPE can manage the traffic load, especially when it might serve as a relay for another CPE. Consequently, the algorithm must account for the CPE's potential role as a network bottleneck. Conversely, in an SD-WAN Core model, this metric focuses on the load of intermediate nodes, such as routers, to prevent congestion and optimize path selection by favoring less congested nodes. The CPE/Node Load metric can be formulated as a normalized value L, defined as:

$$L = w_1 \cdot \frac{U_{\text{CPU}}}{U_{\text{CPU, max}}} + w_2 \cdot \frac{U_{\text{MEM}}}{U_{\text{MEM, max}}} + w_3 \cdot \frac{N_{\text{conn}}}{N_{\text{conn, max}}} \quad (1)$$

where U_{CPU}, U_{MEM}, and N_{conn} represent the current CPU utilization, memory utilization, and number of active connections on a node, respectively, while their denominators denote the maximum capacities. The weights w_1, w_2, and w_3 enable prioritization among these factors based on deployment-specific requirements.

By continuously monitoring the load on CPEs or intermediate nodes, the TE algorithm can make adaptive forwarding decisions. It balances the trade-off between minimizing Net-KPI metrics and mitigating node overload. Additionally, the TE algorithm can dynamically decide whether to utilize an intermediate CPE as a relay based on its current load and network conditions or to establish a direct connection between CPEs, even if it results in higher Net-KPI values, ensuring optimal performance under varying load scenarios.

c) Dependability metrics: Dependability is a qualitative property encompassing reliability, availability, safety, integrity, and maintainability. Reliability ensures correct operation over time, while availability focuses on service readiness. In SD-WAN, these attributes significantly impact TE and routing decisions. Defining failure events in SD-WAN links or tunnels depends on the network designer. A failure may be an inability to transfer traffic for a prolonged period or a violation of Net-KPIs, such as excessive packet loss or latency beyond acceptable thresholds. Failures can also be application-specific, where a tunnel may be unusable for one service but functional for others. Time-to-failure and time-to-repair are modeled as random variables with a probability density function (pdf). Their means are represented as Mean Time To Failure (MTTF) and Mean Time To Repair (MTTR), with Mean Time Between Failures (MTBF) = MTTF + MTTR. These values are not

TABLE V: Load balancing and Dependability metrics for SD-WAN

Metric	Formulation	Description
	Load Balancing Metrics	
Link Utilization Cost	= Cost for keeping the link active (Fixed Cost) + Bandwidth utilization (Variable Cost)	Measures the utilization cost to assess congestion and optimize resource usage.
CPE/Node Load	$= w_1 \cdot \frac{U_{\text{CPU}}}{U_{\text{CPU, max}}} + w_2 \cdot \frac{U_{\text{MEM}}}{U_{\text{MEM, max}}} + w_3 \cdot \frac{N_{\text{conn}}}{N_{\text{conn, max}}}$	Evaluates the current load on a node based on resource utilization and active connections.
	Dependability Metrics	
Tunnel/Link Reliability	$\approx e^{-t/\text{MTBF}}$	Probability of no failures in the link/tunnel over a time interval of duration t.
Tunnel/Link Availability	$= \frac{\text{MTTF}}{\text{MTTF+MTTR}}$	Proportion of time that an SD-WAN link/tunnel is not affected by failures.
Link Redundancy	= Number of backup links	Evaluates the presence of backup links available for activation in case of primary link failures.
Failover Time	= Instant to activate backup − Instant of failure detection	Measures the time it takes for traffic to be rerouted from a failed link to an available backup link.
Site Availability	= 1 − (Failure probabilities of all tunnels multiplied together)	Calculates the probability that at least one WAN tunnel is operational in a branch.

predetermined but can be estimated by analyzing historical monitoring data, assuming a constant rate over time. For instance, MTTF is calculated by summing time intervals between failures and dividing by the total failures, while MTTR is estimated similarly for repairs.

Mathematically, **Reliability**, defined as follows: $R(t) = P\{X > t\} = \int_t^\infty f(x)dx$, is the probability of no failures over a specific time interval $(0, t)$; where X is the random variable representing the time to failure, and $f(t)$ is its probability density function.

For practical applications, reliability is often estimated assuming exponentially distributed failure times as follows: $R(t) \approx e^{(-\lambda t)}$, with a constant rate of $\lambda = MTBF^{-1}$.

This can be useful for making routing decisions, favoring highly reliable tunnels or links to carry traffic characterized by long sessions, such as VoIP, in order to minimize the likelihood of shifting to another tunnel during the session. Conversely, less reliable tunnels can be preferred for short-lived traffic, such as web browsing.

Availability, on the other hand, is the proportion of time that an SD-WAN link or tunnel is operational and available for use at a particular instant as defined in the following formula: $A = \frac{\text{MTTF}}{\text{MTTF+MTTR}}$. Availability takes into consideration both the failure process and the subsequent restoration or repair process. For instance, a link or tunnel with 99.9% availability, also referred to as three-nines, over the course of a year corresponds to a downtime of approximately 8.76 hours within that year. While this formula often applies to individual components, evaluating the end-to-end service availability between two CPEs requires a more holistic view. A framework in [154] calculates this service-oriented availability by modeling dependencies among key system elements like CPEs, tunnels, and the control plane.

In particular, for computing the reliability, we refer to the event that the system is operating continuously without failures in a given interval, while for computing availability, we refer to the event that the system is operating at a given point in time, irrespective of prior failures and repairs. SD-WAN controllers monitor the availability of tunnels and make load-balancing decisions based on the tunnel's current state. Unavailable or degraded tunnels are temporarily avoided until they are back to an operational state. Load balancing algorithms may prefer less susceptible tunnels to disruptions [94], ensuring continued connectivity in the event of link failures.

Other dependability metrics to make informed decisions about network management can be: 1) **Link Redundancy**, is the presence of backup or redundant links, and it can be evaluated by identifying and counting the number of backup links available for activation in case of primary link failures. 2) **Failover Time**, measures the instant it takes for traffic to be rerouted from a failed link to an available backup link. It is measured by recording the duration between the failure of a primary link and the successful rerouting of traffic to a backup link. 3) **Site Availability**, is essential for maintaining reliable network performance. When multiple WAN tunnels are used simultaneously, the overall availability of the site improves. It is determined by calculating the likelihood that at least one tunnel is operational. This is done by considering the failure probability of each tunnel, which is the chance that a specific tunnel will not function properly. The combined probability of all tunnels failing is calculated by multiplying their individual failure probabilities, assuming that they are statistically independent. Finally, the total site availability is calculated as one minus the combined failure probability [155] [156]. This result gives the probability that at least one tunnel remains active, ensuring the site's overall connectivity.

TE algorithms may prioritize the use of resilient links for critical applications to ensure uninterrupted service. TE algorithms use availability percentages to evaluate the dependability of links and make informed decisions about traffic routing [148], [157], [158]. By incorporating dependability measures as performance metrics, TE algorithms can proactively plan for link failures, optimize traffic rerouting, and ensure that network resources are utilized efficiently. To conclude, the Load balancing and Dependability metrics for SD-WAN are summarized in Table V.

3) Policy-based routing and optimization algorithms: SD-WAN administrators define traffic routing policies based on application requirements, business priorities, and WAN traffic types, which fall into three categories: interactive, elastic, and background traffic [109]. Interactive traffic has the highest priority and requires immediate transmission, while elastic traffic consists of bulk transfers with specific deadlines. Background traffic follows a best-effort model with no strict deadlines. In Cloud SD-WANs, interactive traffic represents 5–15% of total traffic, whereas bulk transfers constitute a significant share [109]. The distribution of these traffic types varies depending on the SD-WAN architecture (Core vs. Edge) and application use case, influencing TE methodologies. Overlay network capacities fluctuate over time, making instantaneous measurements unreliable. For instance, tunnels using broadband-based underlay networks experience performance shifts due to variations in the underlying infrastructure. Consequently, TE algorithms must account for uncertainty, adding complexity to path selection. Effective TE algorithms dynamically compute paths based on real-time network conditions and QoS-based policies, recalculating routes when degradation or failures occur. By considering multiple available paths, they enhance redundancy, reliability, and resource utilization. Optimization objectives may include minimizing delay, maximizing bandwidth, or reducing costs. The following sections explore heuristic TE approaches for dynamic path selection and load balancing, followed by ML-based methods that enable intelligent, predictive routing to further optimize SD-WAN traffic management.

a) Heuristic-based algorithms: Heuristic-based algorithms are essential in SD-WAN TE because they effectively adapt to real-time network fluctuations, including congestion, bandwidth variations, and link or tunnel failures. By employing various TE strategies, these algorithms prioritize performance optimization while minimizing operational costs, making them indispensable for large, distributed SD-WAN environments.

A common approach involves **adaptive path selection**, where network performance metrics such as latency, bandwidth utilization, and packet loss are continuously collected and processed by routing algorithms. Modified versions of the Dijkstra algorithm [159] dynamically adjust traffic routes based on real-time network conditions [100], [117], [160]. Unlike the traditional shortest-path approach, these modified algorithms incorporate multi-metric cost functions that factor in latency, bandwidth availability, reliability, and hop count to ensure optimal path selection. The algorithm continuously gathers network metrics, feeding them into the path selection process. Instead of minimizing hop count alone, the multi-metric cost function assigns a weighted cost to each link, evaluating all possible paths from source to destination. The path with the lowest total cost is selected, ensuring traffic is routed along the most efficient and reliable route. This method enhances fault tolerance by proactively rerouting traffic when link conditions degrade, reducing service disruptions [89], [92], [161]. Dynamic rerouting mechanisms allow SD-WAN to switch traffic to backup paths in response to congestion or link failures [99], [160]. This capability is crucial in large, geographically distributed SD-WANs where link reliability varies. By rerouting around degraded links, SD-WAN maintains high-quality service, particularly for latency-sensitive applications like VoIP and video conferencing. These implementations are primarily edge-based, where real-time decisions are made at CPEs. Additionally, application-specific performance indicators, such as QoE metrics from SaaS platforms, can be integrated to further enhance routing efficiency [162].

Another approach is based on **multi-path routing** combined with bandwidth optimization. Multi-path routing distributes traffic across several paths simultaneously, reducing the likelihood of congestion while optimizing the use of available bandwidth. Heuristic algorithms in this category balance traffic loads across multiple paths based on real-time metrics like bandwidth availability and hop count [93], [94], [136]. This not only improves network performance by reducing the number of congested links by up to 50% [92], but also ensures that bandwidth-intensive and latency-sensitive applications receive adequate resources, preventing bottlenecks and improving overall efficiency [100]. Implemented in SD-WAN edge and core models, these solutions are used to optimize traffic flows between enterprise DCs and over multiple WAN links.

The traffic engineering problem can be formulated as an Integer Linear Programming (ILP) optimization at the controller level using the Traffic Matrix (TM) demand. While ILP guarantees optimal solutions, its computational complexity scales significantly with network size. To mitigate this, a partitioned approach divides the TM into smaller subsets, each processed as an independent ILP problem solved in parallel [163]. Unlike naive partitioning, these algorithms optimize flexibility and efficiency, reducing computation time. However, ILP remains impractical for real-time traffic fluctuations, making it more suitable for long-term traffic optimization rather than dynamic adaptation. Beyond ILP-based methods, Fluid Flow Models offer a computationally efficient alternative for analyzing SD-WAN traffic dynamics [164]. These models approximate network congestion and routing behavior using continuous functions instead of discrete packet simulations, enabling rapid evaluation of queue management and multi-path routing strategies. This approach provides a scalable solution for assessing QoS mechanisms in large-scale SD-WAN deployments.

Given the benefits of multi-path routing, threshold-based routing [165] addresses challenges in highly varying traffic load environments, where frequent controller-to-switch synchronization can degrade performance. By dynamically adjusting routes only when predefined thresholds are exceeded, this approach can reduce the synchronization overhead by up to 80% while maintaining up to 87% of optimal load-balancing performance.

Distributed load balancing enhances SD-WAN efficiency by dynamically adjusting traffic routing at the network edge. Solutions like [133] distribute traffic across multiple paths using real-time link utilization data, iteratively refining split ratios until an optimal balance is achieved. This distributed approach improves maximum link utilization, mitigates congestion, and offers a scalable alternative to centralized traffic engineering. Furthermore, integrating MPTCP with SD-WAN enables a single flow to be split into multiple subflows,

dynamically adjusting traffic distribution based on congestion and link failures [132]. These edge-based load-balancing algorithms enhance responsiveness and scalability without strict centralized control. This method is particularly effective for managing large-scale bulk transfers between DCs, where scheduling transfers with deadline awareness is crucial. Heuristic algorithms such as Bulk Transfer Scheduling with Deadline (BESD) optimize best-effort SD-WAN tunnels, adjusting for fluctuating Internet tunnel capacities [115], [161]. Using stochastic optimization, these models predict bandwidth bottlenecks and proactively schedule transfers, improving completion rates and increasing profitability by up to 60% [151]. The objective function accounts for key performance metrics like delay, packet loss, and operational costs, enabling SD-WAN to reroute traffic through less congested paths before network degradation occurs. A real-world example of this approach is Google's B4 WAN [108], where a centralized traffic engineering solution dynamically splits application flows across multiple paths, achieving near 100% link utilization. This architecture balances capacity with application demand, leveraging proactive bandwidth reallocation and predictive traffic adjustments to optimize large-scale network performance [146], [150].

Another key strategy in SD-WAN TE is **intent-based routing optimization**, which aligns traffic routing decisions with business objectives such as minimizing congestion, maximizing QoS, or reducing operational costs [153]. These algorithms predict traffic demands and network conditions, dynamically adjusting edge CPE policies to meet SLAs while optimizing overall network performance. This flexible model prioritizes high-priority applications by allocating sufficient bandwidth, while lower-priority traffic is balanced accordingly. Rate limits and load-balancing policies are continuously adapted based on real-time application requirements and network measurements [150]. A centralized controller monitors overlay link quality and adjusts traffic flows to maintain application-level QoS parameters, such as latency and jitter. By estimating available bandwidth in real-time, intent-based TE algorithms optimize resource allocation and improve SLA compliance.

The discussed heuristic TE algorithms, including adaptive path selection, multi-path routing, distributed load balancing, and intent-based routing, are integral to SD-WAN deployments, operating effectively at both the edge and core levels. At the edge, client organizations leverage these algorithms to dynamically manage traffic between branch offices, DCs, and cloud services. Multi-path routing and bandwidth optimization ensure that critical applications receive necessary resources, while cost-effective traffic distribution across broadband, 4G/LTE, and MPLS links enhances performance. At the core, SD-WAN providers optimize backbone networks using stochastic models and fault-tolerant routing to manage vast geographical traffic loads and ensure service reliability. These providers employ TE algorithms to prevent congestion, reroute traffic during failures, and maintain an efficient, secure, and adaptable overlay network layer for diverse customer needs.

b) ML-based algorithms: After exploring heuristic TE approaches, we now focus on ML-based algorithms, which enhance SD-WAN by enabling predictive, data-driven traffic optimization. Unlike heuristic methods, ML techniques continuously learn from historical and real-time data to anticipate network conditions, making routing more intelligent and adaptive.

In inter-cloud communications, optimizing traffic distribution across links with varying pricing models is crucial for cost reduction. Cost-aware algorithms like Dynamic Traffic Management (DTM) [136] and Grandet [134] use **prediction-based techniques** to estimate traffic volumes and dynamically adjust routing. DTM minimizes operational costs by redistributing manageable traffic, achieving up to 29% savings. Grandet, leveraging the Bootstrap method with neural networks, estimates traffic demands without prior knowledge, improving cost-efficiency under uncertain network conditions. The Bootstrap method is a statistical technique that estimates the distribution of a dataset by resampling it, i.e., random samples are repeatedly drawn with replacement to create multiple new datasets (bootstrap samples). It is widely used to assess the accuracy (like standard error and confidence intervals) of estimates or predictions, especially in cases where traditional assumptions about the data (such as normal distribution) do not hold. In the context of SD-WAN, the Bootstrap method helps estimate the intervals of network parameters (like flow sizes and link qualities) based on previous data. Grandet uses it in combination with neural networks to quantify the uncertainty of network traffic demands. Resampling historical traffic data generates interval estimates with high confidence, which guide cost-effective traffic scheduling decisions even under unpredictable conditions.

Meanwhile, **Reinforcement Learning (RL)** has gained traction in SD-WAN TE due to its ability to adapt dynamically to changing network conditions. RL excels at making decisions in environments where conditions vary in real-time, which is ideal for networks such as SD-WANs with fluctuating traffic, link failures, and bandwidth constraints. RL operates based on the interaction between an agent and an environment. In the context of SD-WAN, the agent (e.g., SD-WAN controller) makes routing decisions, and the environment is the network, providing feedback through rewards or penalties based on network performance. In other words, RL can predict network performance degradation and reroute traffic preemptively to optimize the network's efficiency.

When designing an RL algorithm for SD-WAN, it is important to account for the wide range of actions the agent can take within the network environment, each corresponding to different optimization tasks. The SD-WAN agent can optimize traffic distribution to the edge across multiple overlay tunnels at the CPE level, dynamically adapting to real-time network fluctuations such as changes in delay and bandwidth availability. By adjusting paths in response to these changes, the agent ensures efficient traffic routing, maximizing resource utilization, and maintaining high-quality service.

In this context, **Q-learning** is employed to enable the agent to learn an optimal policy for routing traffic through the network [124]. Q-learning is a model-free RL algorithm that iteratively improves by learning a Q-value function, which approximates the expected cumulative reward for each state-action pair. This helps the agent estimate which actions will

lead to the best long-term outcomes, optimizing traffic flow and reducing QoS policy violations by 33%. By learning from real-time feedback, the Q-learning agent continually refines its routing strategy to achieve long-term performance gains.

In traditional Q-learning, a Q-table is maintained to store the expected rewards for each state-action pair. However, for large state spaces, as in large SD-WAN environments, the Q-function is typically approximated using a **Deep Q-Network (DQN)**, which employs a neural network to estimate Q-values. Specifically, DQN replaces the traditional Q-table with a neural network to approximate the Q-value function, predicting the expected future reward for taking specific actions in a given state. An effective approach to optimize traffic distribution to the edge using DQN is to collect end-to-end delay measurements between CPEs in the SD-WAN through network tomography techniques [97], [98]. These measurements act as feedback for the agent, providing insights into the network's performance without requiring full visibility into the internal structure of the network. Ultimately, the agent learns to optimize traffic distribution based on these tomography-inferred delays by selecting actions that maximize the Q-value and minimize network latency. DQN can also be used to maximize service availability in SD-WAN environments by reducing QoS violations by 33% while optimizing latency and service uptime [157], [158]. A more advanced approach is to collect not only the network delay but also consider the available bandwidth of each tunnel [152].

While DQN excels in discrete-action environments, more complex SD-WAN scenarios, where actions like traffic allocation must be continuously adjusted, require algorithms capable of handling continuous-action spaces. This is where the Soft Actor-Critic (SAC) algorithm comes into play. SAC is designed specifically for continuous-action scenarios and offers advantages such as entropy maximization, which encourages the agent to explore more the action space by preventing premature convergence to suboptimal solutions [147]. However, SAC's adaptability in complex SD-WANs also comes with notable challenges in terms of computational overhead, which can impact deployment costs and latency, and memory resources, potentially complicating scalability in highly dynamic SD-WAN environments with continuous data flows.

Multi-Agent Reinforcement Learning (MARL) is an emerging approach for decentralized TE in multi-site SD-WAN deployments. MARL enables independent agents at each site to optimize overlay selection based on localized metrics, reducing control overhead and improving scalability [149], [166], [167]. However, coordination between agents remains a challenge, as local optimizations may conflict with global network objectives. To optimize SD-WAN controller load balancing, Deep Deterministic Policy Gradient (DDPG) has been introduced for continuous traffic migration decisions [168]. By dynamically redistributing traffic across controllers based on real-time metrics, DDPG effectively reduces migration costs and enhances network stability compared to traditional RL methods. In summary, RL enhances SD-WAN TE by proactively optimizing multiple objectives, such as latency, load balancing, and cost reduction. However, its practical deployment faces scalability challenges due to large state-action spaces and computational overhead. Designing effective reward functions is critical to guiding RL agents toward optimal performance. Table VI summarizes key research contributions in TE for SD-WAN.

4) Emerging challenges and research directions in SD-WAN TE: As SD-WAN continues to evolve, addressing the increasing demand for scalable, flexible, and resilient network solutions, novel approaches to TE are necessary. While existing TE methods enable more efficient traffic management and application-aware routing, several emerging research areas seek to enhance the adaptability, scalability, and security of TE within SD-WAN environments.

Integrating **Segment Routing (SR)**, especially SRv6, into SD-WAN is gaining traction due to its potential to offer fine-grained, policy-based routing capabilities [104]–[106]. Incorporating SRv6 into SD-WAN can facilitate multi-path routing that dynamically adapts to current network conditions. Through segment-based path selection, SRv6 enables traffic to be routed across multiple paths simultaneously, optimizing resource utilization and balancing load more effectively across network links. Moreover, SRv6's architecture supports large-scale networks by reducing the complexity of maintaining state information at intermediate nodes. Another relevant point is that SRv6 supports service function chaining [105], which enables the orchestration of multiple services (like firewalls, load balancers, and other virtual network functions) in a single data path. In this context, the CPEs act as gateways between customer sites and classify traffic into specific end-to-end network slices. Each slice is essentially a Layer 3 VPN (L3VPN) running over a WAN, isolated from other slices. Each L3VPN is dedicated to a specific group of users or services. The traffic associated with each slice is forwarded between customer sites and the cloud through CPEs. This capability is crucial for deploying value-added services efficiently within the SD-WAN framework.

A major challenge in implementing SRv6 within SD-WAN is the header overhead associated with using long IPv6 addresses, which increases packet size. This can impact performance, especially in high-traffic environments where efficient bandwidth utilization is critical. Emerging research is exploring header compression techniques [103], such as Hop Compression based on Multi-Dimensional Routing Differences (HCMD) [104], which aim to reduce the size of SRv6 headers by minimizing the number of required intermediate hops. This approach not only saves bandwidth but also improves link utilization and enables SRv6 to scale more effectively in large networks. Further advancements in SRv6 compression methods will be critical in making SRv6 a viable option for large-scale SD-WAN deployments.

Future research will focus on developing algorithms that leverage SRv6 multi-path capabilities into SD-WANs to balance traffic loads in real-time and dynamically re-routing traffic to avoid congestion. The goal is to enhance QoS for latency-sensitive applications and increase overall network efficiency, even under fluctuating traffic patterns and network failures.

While heuristic methods provide quick responses to network

TABLE VI: Summary on Traffic Engineering algorithms

Traffic Engineering		
Traffic Classification	Heuristic-based Traffic Classification	[93], [122], [132]
	Machine-Learning-based Classification	[141]
Performance Metrics	Network Performance Indicators	Throughput [124], [132], [133], [136], [146]
		Delay [89], [97], [115], [117], [147]–[149]
		Packet Loss [99], [150], [151]
		Bandwidth [93], [98], [108], [110], [134], [146]
	Load Balancing Metrics	Link utilization, cost, and capacity [10], [100], [153]
		Path Availability [94]
	Service Availability	[148], [157], [158]
Policy-based Routing and Optimization	Heuristic-based Algorithms	[89], [92]–[94], [99], [100], [108], [115], [117], [132], [133], [146], [150], [151], [153], [160], [161]
	ML-based Algorithms	[97], [98], [124], [134], [136], [147], [149], [157], [158], [166], [168]

changes, they lack predictive capabilities. ML offers predictive and adaptive potential, but it requires substantial computational resources. Research is shifting toward **hybrid TE approaches**, combining the simplicity of heuristic algorithms with the adaptability of ML and RL models. For example, heuristic methods could handle immediate routing adjustments, while ML models predict traffic patterns based on historical data to provide insights for future optimization. Additionally, RL can continuously learn from real-time network conditions to adjust traffic routing proactively, maximizing QoS for latency-sensitive applications. Moreover, as SD-WAN deployments scale, centralized TE approaches become increasingly resource-intensive.

Multi-agent reinforcement learning offers a decentralized alternative where each network site can make localized routing decisions. These agents can optimize path selection independently while sharing performance data with other agents to ensure global TE goals are met. This decentralized framework enhances scalability by reducing the computational burden on centralized controllers. However, challenges remain in coordinating multiple agents, particularly in balancing localized decisions with network-wide objectives. Hybrid models tailored to specific SD-WAN use cases, such as reducing operational costs or optimizing latency, are a promising area for future exploration.

In SD-WAN, **intent-based routing** is gaining attention as a strategy for aligning TE decisions with high-level business goals, such as minimizing latency, ensuring high availability, or reducing costs. These approaches use policy-driven optimization, where SD-WAN controllers adapt routing strategies based on predefined intents. Emerging research is exploring how SRv6 policies can support intent-based routing, where segments are assigned based on business objectives, and TE algorithms dynamically adjust routing decisions to ensure these objectives are met. Intent-based approaches hold potential for reducing human intervention in traffic management, as they provide automated, policy-aligned TE with minimal configuration requirements [153].

In summary, emerging research directions in TE for SD-WAN aim to create more efficient, adaptive, and secure traffic management frameworks. By leveraging multi-path routing, adaptive algorithms, decentralized models, and enhanced monitoring, these approaches will provide SD-WAN architectures with greater flexibility to meet evolving network demands. The development of hybrid and intent-based methods, as well as the incorporation of SRv6 into TE, will be critical as SD-WAN environments continue to scale and diversify in complexity.

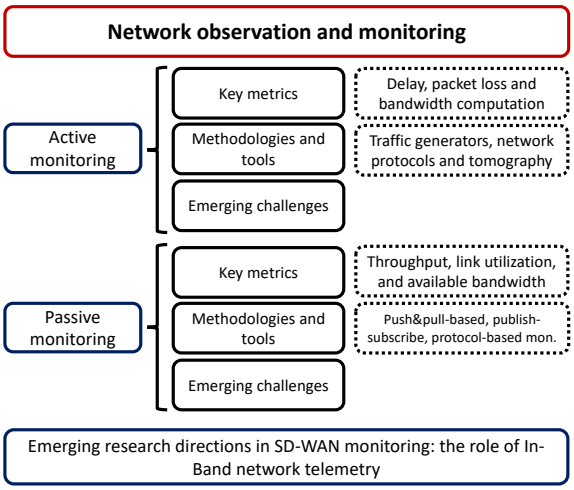

Fig. 10: Taxonomy of network observation and monitoring.

B. *Network observation and monitoring*

SD-WAN represents a dynamic and flexible approach to WAN, leveraging SDN principles to optimize network traffic, enhance application performance, and simplify network management. In this context, effective monitoring is essential for gaining insights, detecting anomalies, and addressing issues promptly to ensure network reliability and performance. Several frameworks have been developed within the SDN ecosystem to facilitate network monitoring. Examples include Yaz [169], OpenTM [170], MicroTE [171], OpenSketch [172], PayLess [173] and OpenSample [174]. Each of these frameworks is designed to retrieve and process SDN network performance metrics effectively, providing insights into network status and enabling quick response to performance issues.

However, SDN has been primarily deployed within DCs or localized network environments, as such, SDN monitoring focuses on managing and optimizing network resources in a controlled, often homogeneous setting. Monitoring in SDN typically targets elements within a single administrative domain, such as switches and VNFs. On the other hand, SD-WAN operates across wide geographic areas, often involving multiple types of network connections (e.g., MPLS, broadband, 4G/LTE). SD-WAN monitoring must account for variable-quality network links and more dynamic, heterogeneous environments across different administrative domains, such as branch offices, cloud services, and DCs.

As shown in Fig. 10, we can identify different monitoring approaches that are categorized as active and passive. They offer distinct advantages and challenges, each suited to particular performance insights and operational needs. **Active monitoring** involves the injection of probe traffic to measure network performance in real time, making it ideal for proactive diagnostics and SLA compliance checks. **Passive monitoring**, on the other hand, observes actual traffic patterns without adding load, offering a realistic view of user and application behavior, which is valuable for long-term performance assessment and security analysis. Moreover, **In-band Network Telemetry (INT)** emerges as a promising technique for SD-WAN monitoring. INT integrates telemetry data into the network packets themselves, enabling real-time tracking of network state without requiring additional probe traffic. This approach combines the strengths of active and passive monitoring by providing granular, continuous visibility into network performance while minimizing overhead and avoiding disruption to regular traffic flows. Within this survey, we address the potential of in-band network telemetry as a future research direction in Section VI-B3, where its capabilities for dynamic SD-WAN environments are explored.

1) Active monitoring: Active monitoring is a key component of SD-WAN architecture, enabling real-time performance evaluation, proactive traffic management, and SLA compliance. By injecting probe packets, this approach gathers critical metrics such as delay, jitter, packet loss, and bandwidth availability, allowing SD-WAN controllers to make dynamic, data-driven routing decisions. For applications like VoIP and video conferencing, real-time monitoring ensures consistent quality by detecting link degradation and congestion before they impact service availability. This section explores methodologies and tools such as Bidirectional Forwarding Detection (BFD) [175] and STAMP [176], highlighting their role in achieving high-performance SD-WAN operations. Additionally, it examines trade-offs, including resource overhead and potential traffic interference, while discussing AI-driven monitoring techniques designed to enhance scalability, efficiency, and adaptability in complex SD-WAN environments.

a) Key metrics: Among the various performance monitoring metrics, **delay**, **packet loss**, and **bandwidth** estimation play a central role in SD-WAN because they directly impact the QoS for critical applications and are essential for effective traffic management.

In SD-WAN, we can identify different types of **delay**. The first is called Control Plane Delay (CPD) and is defined as the end-to-end delay between network nodes (or CPEs) and the controller. Since the controller is at a different distance from the nodes, the CPD affects the performance of control functions, such as instructions dictated by TE or real-time monitoring algorithms. The second metric of interest is referred to as Overlay Delay (OD). It quantifies the time it takes for a packet to traverse the physical infrastructure of the various underlay networks that support the virtual overlay network. In the context of SD-WAN Edge scenarios, the underlay network operates as a self-contained entity, often perceived as a black box. Consequently, obtaining insights into its internal workings can be challenging. In such cases, the primary avenue for gaining knowledge about the underlay network is through end-to-end tomography measurements, particularly those related to average delays [97], [160]. The third is called Node Processing and Transmission Delay (NPTD) and consists of the time required to push all the bits of a data packet onto the overlay network link. Depending on the type of node and CPE, data packets may need to be processed by different software modules, such as the security module, encryption module, or even other applications in the different VMs (or containers) that compose the CPE when they are sent. Therefore, there may be a significant increase in the transmission delay that must be taken into consideration [177] [178].

Packet loss monitoring is critical for maintaining service reliability and quality, especially for applications requiring data integrity and consistency, such as financial transactions, remote diagnostics in healthcare, and collaborative SaaS platforms. SD-WAN's centralized management and real-time monitoring capabilities allow continuous evaluation of packet loss across diverse underlay networks, including broadband, 4G/LTE, or MPLS. If significant packet loss is detected on one path, SD-WAN controllers can instantly switch traffic to a more reliable link, leveraging the overlay's dynamic routing capabilities [117].

Efficient **bandwidth monitoring** is essential for optimizing network performance and ensuring that critical applications receive the necessary resources [179]. SD-WAN leverages overlay tunnels, a key component of its architecture, to enhance network agility and flexibility. However, since the overlay tunnels are often built over one or more underlay public networks, characteristics like bandwidth or delay fluctuate over time because other connections in the same networks behave unpredictably. Monitoring bandwidth within these overlay tunnels is crucial for maintaining network quality and providing SLAs for SD-WAN services.

In the next section, we will explore the methodologies used to measure these critical performance metrics, laying the groundwork for a deeper understanding of their practical applications and impact on SD-WAN networks.

b) Methodologies and tools: Active monitoring in SD-WAN relies on **probing techniques**, where synthetic packets are injected into the network to measure key performance metrics such as delay, jitter, packet loss, and bandwidth availability. These methods optimize traffic management, ensure SLA compliance, and enable real-time network adaptability. Probing techniques provide granular insights into processing delays within CPEs by timestamping packets as they pass through firewalls, encryption modules, and virtualized applications [177], [178]. This allows administrators to pinpoint bottlenecks, optimize resource allocation, and assess trade-offs between functionality and latency in complex SD-WAN environments.

Traffic generator tools such as **iPerf** [180], **D-ITG** [181], and **Ostinato** [182] create synthetic traffic to simulate real-world conditions. These tools enable real-time monitoring of end-to-end performance by evaluating key metrics under varying network loads. D-ITG has been used to implement monitoring software agents within CPEs to measure packet

delay, jitter, and loss. These metrics are transmitted to the SD-WAN controller's central monitoring module [117], [183]. Active overlay tunnels generate monitoring flows between CPEs, with the receiving CPE calculating delay and packet loss before forwarding the data to the active monitoring module. This module, operating periodically every *T* seconds, analyzes the collected data to ensure compliance with business policies. If deviations are detected, corrective actions are initiated to maintain service quality and optimize network performance.

In addition to synthetic traffic generation, SD-WAN utilizes **network protocols** such as **ICMP, BFD, STAMP**, and **BGP** to assess performance metrics. **ICMP** is widely used for basic connectivity tests and delay measurement via Ping. While effective for quick diagnostics, ICMP packets are often deprioritized or filtered, limiting their accuracy in complex SD-WAN environments. **BFD** [184], [185] provides high-frequency monitoring for link liveliness, ensuring reliable paths with detection intervals as low as *50* ms. This rapid response capability enables SD-WAN controllers to reroute traffic dynamically. Recent research [186] employs DQN to optimize BFD probe frequency, improving throughput and resource efficiency. **STAMP** [176] enhances monitoring by enabling timestamp-based measurements for one-way and round-trip delay analysis. It is particularly useful for evaluating Segment Routing (SRv6) paths [106], [187], offering detailed insights into latency and packet loss. STAMP's compatibility with software-defined environments facilitates seamless integration into open-source SD-WAN solutions. **BGP** contributes to SD-WAN monitoring by tracking dynamic routing changes and ensuring path stability across WAN connections [188]. It detects route flapping and enables proactive traffic rerouting. In overlay networks, BGP updates help assess underlay path availability, allowing SD-WAN controllers to redirect traffic when performance degradation is detected.

Probing techniques also support **network tomography**, inferring hidden network characteristics such as congestion points and underlay link quality [189]–[191]. By analyzing delay correlations between overlay paths, SD-WAN can detect shared underlay congestion and trigger topology reconfiguration while minimizing disruptions [97], [100]. In this context, **bandwidth estimation** [179] plays a crucial role in optimizing SD-WAN resource allocation. Techniques such as TOPP [192], PathLoad [193], and PathChirp [194] provide accurate assessments of available bandwidth. These predictions guide SD-WAN orchestration systems, ensuring efficient resource provisioning. By combining active probing, traffic generation, and network tomography, SD-WAN enhances its ability to dynamically adapt to changing network conditions, ensuring high performance and SLA compliance. Table VII summarizes the advantages, drawbacks, and challenges of the active probing techniques investigated so far.

c) Emerging challenges: Active monitoring in SD-WAN is essential for maintaining network performance, ensuring SLA compliance, and optimizing resource allocation. However, as SD-WAN environments scale, challenges related to scalability, resource efficiency, AI/ML integration, bandwidth estimation, and security must be addressed.

Scalability. Expanding SD-WAN deployments to hundreds or thousands of nodes introduces challenges in handling the overhead from probing-based methodologies such as BFD. High-frequency probing (e.g., BFD at 50 ms intervals) can congest network links, particularly in public Internet-based SD-WANs [184]. A key design principle recommends keeping data plane monitoring overhead below 5% of total capacity [195]. Solutions include dynamic probing mechanisms that adjust probe frequency based on real-time conditions and distributed monitoring architectures to reduce controller load, though these introduce synchronization challenges. Efficient data aggregation techniques are also needed to compress monitoring data while preserving critical insights.

Minimizing Resource Overhead. Active monitoring increases computational demands, particularly in timestamp-based delay measurements within CPEs, where packets pass through multiple processing stages (e.g., firewalls, encryption, IDPS) [177], [178]. Monitoring tools such as iPerf and D-ITG further consume bandwidth and processing resources, leading to potential performance degradation in large deployments. Lightweight protocols like STAMP reduce monitoring impact, and integrating **hardware acceleration** (e.g., hardware-based timestamping) could significantly lower computational overhead.

Integration with AI/ML for Predictive Analytics. AI and ML offer predictive monitoring capabilities, enabling congestion and packet loss forecasting [186]. However, SD-WAN's reliance on opaque underlay networks limits data availability, affecting model accuracy. Additionally, ML models require extensive computational resources, frequent retraining, and seamless integration with real-time monitoring. MARL could decentralize decision-making, reducing the strain on centralized controllers, but coordinating multiple agents while ensuring network-wide optimization remains a challenge.

Advancing Bandwidth Estimation Techniques. Accurate bandwidth estimation is critical for SD-WAN resource optimization, particularly in environments with fluctuating congestion. Traditional techniques such as PathLoad and PathChirp [193], [194] struggle to account for cross-traffic interference. Advanced statistical models and ML-based approaches could enhance estimation accuracy, allowing SD-WAN controllers to predict bandwidth availability and dynamically adjust resource allocations.

Security in Active Monitoring. Monitoring traffic, including probe injections and timestamping, poses security risks, particularly in multi-tenant SD-WAN environments. Attackers could intercept monitoring packets to infer network topology and exploit vulnerabilities. Encrypting probe traffic prevents unauthorized access, while isolating monitoring data within dedicated logical channels enhances security. AI-driven anomaly detection can identify abnormal monitoring patterns indicative of security threats, though real-time processing constraints in edge environments present challenges. Balancing security and performance will be a key focus for future research.

Addressing these emerging challenges will be crucial for ensuring SD-WAN's ability to scale while maintaining high performance, efficiency, and security.

TABLE VII: Active probing tools and approaches

Tool / Methodology	Advantages	Drawbacks	Challenges
iPerf / iPerf3	- Effective for testing maximum achievable bandwidth and throughput - Provides comprehensive metrics for bandwidth, packet loss, and jitter	- High traffic volume generated can congest the network - Typically used for scheduled tests rather than continuous monitoring	- Needs dedicated time windows to avoid interference with production traffic - Heavy resource usage during tests; unsuitable for constant real-time monitoring in production
D-ITG	- Granular control over traffic patterns, enabling detailed simulation of various network conditions - Can emulate specific traffic types, making it useful for performance testing of diverse applications	- Generates significant traffic that can impact network bandwidth - Can impact latency-sensitive applications if run continuously	- Requires significant processing resources, particularly if simulating high-load scenarios - Scheduling tests without impacting normal traffic flow in real-time scenarios
Ostinato	- Supports multi-protocol traffic generation, providing insights across different network layers - Allows customized packet streams for end-to-end testing across protocols	- Complex setup and higher CPU usage on edge devices - Can increase latency and jitter for other network traffic during tests	- Managing resource utilization in network environments with limited capacity - Difficult to balance probe complexity with the need for efficient monitoring
ICMP	- Simple and low-overhead method for assessing latency and reachability - Lightweight and quick to implement	- Limited to basic latency and connectivity - Not suited for detailed performance analysis	- Often blocked by firewalls; limited diagnostic capability in complex networks - Provides limited granularity and lacks insights into traffic variability
BFD	- High-frequency probing detects link failures nearly instantaneously, enhancing SLA compliance - Rapid response to link issues ensures reliable path selection	- High probe frequency can lead to network overhead and impact performance under congestion - May cause jitter, especially on low-bandwidth links	- Requires careful tuning of probe intervals to avoid impacting other critical traffic - Processing demands on edge devices, which can affect CPU utilization on resource-constrained devices
STAMP	- Enables precise, timestamp-based delay monitoring, especially effective for SRv6-based networks - Minimal impact on data traffic with smaller probe packets	- Requires time synchronization, which may add setup complexity - Limited support for certain edge devices without compatible firmware	- May require specialized hardware for highly precise monitoring (e.g., time-synchronized clocks) - Needs support for SRv6 in network infrastructure for effective deployment
BGP	- Provides insights into route stability, critical for path availability monitoring in multi-ISP environments - Assists with proactive path adjustments for resilient routing	- Limited to routing-level visibility, lacks application-layer metrics - Route flapping or frequent BGP updates can skew monitoring results	- Balancing probe frequency for timely data without overloading routing infrastructure - Requires integration with SD-WAN controller to interpret BGP route changes dynamically
Network Tomography	- Enables predictive capabilities for congestion and load distribution based on historical network data - Optimizes path selection by forecasting potential bottlenecks	- High computational demands; requires large datasets for accuracy - Complexity of deployment and ongoing maintenance requirements	- Continuous model training needed to ensure accuracy in dynamic traffic environments - Risk of inaccurate predictions under sudden, unforeseen network conditions

2) Passive monitoring: Passive monitoring provides a non-intrusive method for collecting network data by analyzing existing traffic flows without injecting additional packets. Unlike active monitoring, which generates synthetic traffic, passive monitoring observes real-time network activity, making it ideal for SD-WAN environments where efficiency and minimal disruption are priorities. Traditional network settings rely on "capture-and-analyze" tools that require specialized instrumentation. For example, IPMON [196] captures TCP/IP packet headers, timestamps them, and forwards the data for centralized analysis. In SDN-based networks, the OpenFlow [72] protocol retrieves flow statistics directly from switch flow tables. FlowSense [197] utilizes OpenFlow control messages, such as `FlowRemoved` messages, which provide insights into link utilization by reporting flow duration, transmitted bytes, and input port data. However, FlowSense's accuracy depends on flow duration, making it most effective in environments with numerous short-lived flows, such as DC networks. In SD-WAN, where network topologies are more complex with long-lived traffic flows, passive monitoring must be designed for scalability without introducing additional processing overhead.

This section explores key methodologies and tools, including OpenFlow [72] and the extended Berkeley Packet Filter (eBPF) [198], highlighting their role in passive SD-WAN monitoring.

a) Key metrics: Passive monitoring captures key performance metrics such as throughput, link utilization, and available bandwidth (AWB) by leveraging existing traffic flows. Since it does not inject synthetic traffic, passive techniques are lightweight, ensuring minimal resource overhead while delivering actionable insights. Metrics such as packet retransmission rates, congestion window sizes, and RTT can be extracted from TCP flows, enabling visibility into performance trends and potential bottlenecks without additional network load. Additionally, traffic throughput can be inferred using OpenFlow statistics, allowing real-time analysis of data flows and bandwidth utilization [93], [108], [110]. Despite its non-intrusive nature, passive monitoring can introduce software and hardware overhead, particularly in high-speed SD-WAN environments. Capturing packets on a 100 Gbps link requires specialized hardware, such as network taps or high-performance Network Interface Cards (NICs), to prevent packet loss. Processing such high volumes of traffic in real time demands frameworks like Data Plane Development Kit (DPDK) or P4-enabled switches. Without careful design, the computational burden of extracting and analyzing metrics such as retransmission rates and RTT can create monitoring bottlenecks, impacting scalability. To address these challenges, hardware accelerations, such as Field Programmable Gate Array (FPGA)-based processing, offer efficient solutions for

handling ever-increasing high-speed network demands [199].

Similar to active monitoring, the delay is a critical metric also observed in passive monitoring systems. Here, passive methods can infer delay metrics by capturing timestamps of packets as they traverse various network nodes. For example, Transport-Layer Passive Monitoring (TMP) systems use TCP-level metrics to measure RTT and identify issues like retransmissions caused by congestion or tunnel failures [126], [200]. Passive monitoring also provides valuable insights into long-lived flows, which are common in SD-WAN environments, particularly across overlay VPN tunnels implemented over diverse WAN technologies. Packet loss can be inferred passively by analyzing sequence numbers and retransmission behavior in protocols like TCP. For UDP traffic, passive tools capture flow statistics to estimate loss rates. Similarly, throughput and tunnel utilization are calculated by aggregating statistics such as transmitted bytes and flow durations obtained from network devices. Advanced estimators such as the Silent Available Bandwidth Estimation (SABE) model overlay links as single-server queuing systems, leveraging existing latency and packet loss measurements for available bandwidth estimation [150].

However, passive monitoring has its limitations. It relies on the presence of actual user traffic, meaning its effectiveness can diminish during periods of low utilization or idle states. Metrics such as latency or jitter may be challenging to infer accurately without sufficient real-time traffic. Furthermore, passive methods may struggle to detect silent failures or network partitions, as they do not proactively probe network paths. To address these gaps, passive monitoring is often complemented by active techniques, creating a hybrid approach that balances efficiency with comprehensive coverage. These challenges will be further discussed in the following paragraphs.

b) Methodologies and tools: Passive monitoring captures and analyzes network data as it traverses key elements such as CPEs and end-hosts. Several methodologies, including OpenFlow-based monitoring, publish-subscribe methods, and extended Berkeley Packet Filter (eBPF), enhance SD-WAN passive monitoring capabilities.

OpenFlow-based monitoring operates in push-based and pull-based modes. In **push-based monitoring**, control messages between the controller and switches are analyzed to infer bandwidth utilization without additional overhead [201]. In **pull-based approaches**, the controller explicitly requests flow statistics, such as packet loss, from switches using OpenFlow messages like StatsReq and StatsRes [93], [108], [110]. These statistics, including port, queue, and flow table data, provide valuable insights into congestion points, traffic distribution, and QoS compliance [202]. However, frequent polling can introduce latency and computational overhead, necessitating optimizations such as adaptive polling intervals and sampling techniques.

Publish-subscribe monitoring improves efficiency by using a centralized database, such as Redis [203], to store and track network statistics [89]. This model decouples data producers (e.g., routers) from consumers (e.g., monitoring applications), allowing real-time updates on link failures and performance trends. While it enhances responsiveness, scalability challenges arise if the database server becomes a bottleneck. Addressing these issues requires distributed clusters and fault-tolerant mechanisms.

Protocol-based monitoring focuses on transport-layer metrics for delay-sensitive applications. Transport-Layer Passive Monitoring (TPM) analyzes TCP retransmissions, congestion window sizes, and RTTs to assess network health [126], [200]. By leveraging eBPF [198], TPM embeds Kernel Agents (KAs) in the Linux kernel to capture real-time transport-layer events, such as retransmissions, and processes them in User Agents (UAs). When quality constraints are violated, alerts trigger dynamic routing adjustments at the SD-WAN controller. This method ensures precise monitoring but is limited in scope for UDP-based applications like VoIP and online gaming.

In summary, OpenFlow monitoring provides a network-wide perspective on traffic flows, publish-subscribe frameworks enable real-time event tracking, and TPM offers protocol-level granularity. Integrating these techniques creates a multi-layered monitoring architecture, enhancing SD-WAN agility, reliability, and performance. Future optimizations should focus on reducing computational overhead, improving scalability, and expanding protocol compatibility for broader SD-WAN applications.

Table VIII summarizes the advantages, drawbacks, and challenges of the passive monitoring techniques investigated so far.

c) Emerging challenges: Passive monitoring in SD-WAN is essential for maintaining network performance, reliability, and scalability. However, as SD-WAN architectures grow in complexity, monitoring techniques such as OpenFlow-based monitoring, publish-subscribe frameworks, and protocol-based methods using eBPF face several challenges. Addressing these challenges is critical to advancing passive monitoring while meeting modern network demands.

Scalability remains a major concern as SD-WAN deployments expand. OpenFlow-based monitoring relies on high polling frequencies to gather statistics from switches, which can overwhelm centralized controllers, leading to latency and degraded performance. For instance, frequent polling in large enterprise networks can strain controllers, delaying the detection of link failures or congestion [204]. Publish-subscribe frameworks, while efficient for real-time updates, may struggle with high event surges, leading to bottlenecks in central databases [205]. Protocol-based monitoring using eBPF provides detailed transport-layer insights, but processing numerous kernel-level events in high-flow environments imposes significant computational demands [206]. Addressing these challenges requires distributed architectures that decentralize data collection and processing, reducing controller load. Optimizing data collection methods, such as adaptive polling in OpenFlow or selective subscriptions in publish-subscribe systems, can alleviate bottlenecks. Additionally, integrating scalable machine learning models can enable predictive analytics, reducing the need for constant real-time monitoring.

The absence of **standardization** across passive monitoring techniques presents interoperability challenges. Different vendors use proprietary monitoring tools, creating incompatibilities that hinder seamless multi-vendor deployments [207],

TABLE VIII: Passive monitoring tools and approaches

Tool / Methodology	Advantages	Drawbacks	Challenges
Push-Based OpenFlow Monitoring	- Minimal overhead as it relies on controller-switch communication. - Scalable for large networks.	- Limited granularity in metrics. - Requires integration with OpenFlow-enabled devices.	- May not capture detailed real-time events effectively. - Ensuring synchronization across multiple switches.
Pull-Based OpenFlow Monitoring	- Provides detailed, customizable insights into flow statistics. - Suitable for troubleshooting specific network paths.	- High latency and computational overhead in large deployments. - Can introduce significant delays under high polling frequency.	- Efficient synchronization of polling intervals to minimize redundancy. - Balancing granularity with scalability.
Publish-Subscribe	- Real-time updates on events like link failures and recoveries. - Decouples data producers and consumers for modularity.	- Centralized database can become a bottleneck. - Requires robust fault-tolerance mechanisms for the central database.	- Ensuring scalability in high-volume environments. - Efficient handling of simultaneous subscriptions in distributed systems.
Protocol-Based Monitoring (eBPF)	- Granular insights into transport-layer metrics like retransmissions and RTT. - Low overhead by targeting specific kernel-level events. - Enables real-time detection of performance issues.	- Limited to protocol-specific monitoring (e.g., TCP). - Deployment complexity as it requires agents on Linux-based servers. - Incompatibility with non-Linux systems.	- Extending functionality to support other protocols like UDP. - Optimizing for large-scale traffic flows to maintain real-time responsiveness. - Managing resource demands as the number of monitored flows increases.

[208]. OpenFlow, publish-subscribe models, and eBPF operate independently, making unified monitoring strategies difficult to implement. Vendor-specific tools further restrict integration, forcing organizations into proprietary ecosystems. Establishing common protocols and integration frameworks would enable interoperability, reducing vendor lock-in while facilitating cohesive monitoring across SD-WAN infrastructures.

Security and reliability pose significant challenges in SD-WAN monitoring. OpenFlow-based systems suffer from synchronization delays, affecting real-time fault detection [110]. Publish-subscribe frameworks rely on centralized databases, introducing single points of failure that can disrupt monitoring operations [209]. Protocol-based methods like eBPF require careful kernel-level integration to prevent system instability. An improperly implemented eBPF program, for example, could degrade system performance by consuming excessive CPU resources. Enhancing fault tolerance in centralized frameworks, deploying distributed databases, and rigorously testing eBPF implementations can improve reliability while mitigating security risks.

The growing **complexity** of SD-WAN generates vast monitoring data, necessitating advanced analytics for actionable insights. Processing high volumes of real-time traffic requires substantial computing power, and manual analysis is increasingly inadequate [210]. Adaptive monitoring frameworks using machine learning can dynamically adjust sampling rates based on traffic patterns, ensuring critical traffic is prioritized without overwhelming resources. AI-driven analytics offer predictive maintenance, anomaly detection, and automated network adjustments, but implementation challenges include extensive training data requirements and high computational overhead. Research efforts are focusing on refining AI models for network monitoring to enhance predictive accuracy and reduce infrastructure strain [211].

Addressing these challenges will be essential to ensuring passive monitoring can scale effectively, integrate seamlessly across diverse SD-WAN architectures, and provide real-time, intelligent insights for network optimization.

3) Emerging research directions in SD-WAN monitoring: the role of In-Band network telemetry: Network telemetry has become essential for automating network data collection and monitoring [212]. It involves the remote retrieval and processing of network information to enhance visibility and performance optimization.

INT is a prominent network telemetry approach that has gained significant attention in both academia and industry [213]. Unlike traditional active and passive measurement techniques, INT integrates network state collection and metadata insertion directly into traffic packets. This allows telemetry data and user data to share the same network link or even the same packet, enabling real-time monitoring without requiring intervention from the control plane [214]. The source node embeds telemetry tags or instructions within packets, guiding network devices on the data to be collected. The data plane then populates and forwards these packets, ensuring transparency for end hosts. Upon reaching the final hop, telemetry data is extracted and sent to a telemetry server for analysis. This paradigm shift enables network administrators to capture transient issues at the forwarding plane, improving responsiveness and accuracy in SD-WAN monitoring.

INT can achieve fine-grained network monitoring by encapsulating data plane states into traffic packets. However, as an underlying device-level primitive, INT cannot be directly applied to overlay network monitoring, given that underlay and overlay networks typically operate independently and transparently from one another. INT represents a transformative approach to SD-WAN monitoring by embedding telemetry metadata directly into packets traversing the network. This enables real-time, fine-grained monitoring without the additional overhead of probe traffic, bridging the gap between passive and active monitoring. INT's ability to combine network performance assessment with data transmission within the same packets provides unparalleled efficiency and visibility, making it an essential tool for modern SD-WAN architectures.

In SD-WAN environments, INT must be adapted to function over various tunneling protocols such as GRE, IPsec, and VXLAN [215]. These tunnels connect CPEs across different sites, forming the backbone of SD-WAN connectivity. A basic implementation idea is shown in Fig. 11, where the encapsulated packets are routed between two CPEs representing the tunnel endpoints [216]. Once the data reaches the tunnel endpoint, the metadata information, just like the INT header, is unpacked and processed while the original packet is forwarded to the correct destination.

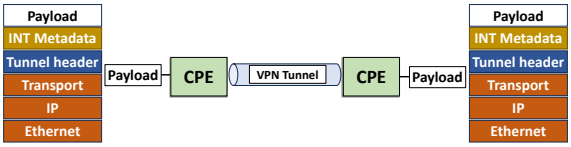

Fig. 11: The principle of the tunnel technology with INT.

There are basic prerequisites in order to let the solution work correctly. The first one is that both network nodes must support INT technology and are configured with the same INT parameters. The second one is that the route between the two endpoints is reachable, and a tunnel connection has been established (e.g., GRE, IPsec, VXLAN tunnel). The described implementation underscores that the telemetry system accomplishes end-to-end state telemetry without affecting the awareness of end users. This telemetry data could be used for multiple essential purposes, including network state measurement, network failure detection, fault localization, and facilitating recovery measures [114].

One of the primary benefits of INT is its ability to provide real-time insights into network performance metrics such as One-Way Delay (OWD) and Two-Way Delay (TWD). The integration of eBPF technology further enhances this capability by allowing telemetry metadata to be dynamically inserted and extracted at the CPE level, ensuring minimal latency and high bandwidth efficiency [217]. Additionally, INT's hybrid monitoring approach reduces the traffic overhead typically associated with active monitoring methods while maintaining the accuracy and responsiveness needed for real-time network management.

Despite its advantages, implementing INT in SD-WAN is not without challenges. One significant concern is the potential impact on network performance due to the increased packet size caused by embedded telemetry metadata. This can strain network throughput and processing capabilities, especially in high-traffic scenarios. Furthermore, ensuring the security and integrity of telemetry data while maintaining line-rate processing poses additional hurdles. For example, embedding telemetry within IPsec-encrypted packets requires careful handling to avoid compromising either security or performance.

Another challenge lies in the interoperability of INT systems across diverse SD-WAN deployments. Different tunneling technologies and configurations may require customized adaptations of INT, increasing the complexity of deployment. Moreover, achieving synchronization between CPEs to accurately measure delays, particularly OWD and TWD, necessitates robust time synchronization mechanisms like NTP, adding to the operational overhead.

To address these challenges, future research should focus on optimizing the integration of INT with tunneling protocols and enhancing the scalability of telemetry systems in SD-WAN environments. For instance, leveraging distributed INT-enabled CPEs can decentralize telemetry processing, reducing the load on central controllers. Additionally, advancements in eBPF technology can further streamline the manipulation of telemetry metadata, ensuring high performance without compromising security.

Exploring new algorithms for compressing telemetry data and minimizing overhead will also be critical. Moreover, combining INT with AI/ML-based analytics can enable predictive monitoring and automated fault resolution, enhancing the overall reliability and efficiency of SD-WAN networks.

In conclusion, INT offers a promising avenue for advancing SD-WAN monitoring, providing real-time insights, and improving network visibility. However, its successful implementation requires overcoming challenges related to performance, interoperability, and scalability, which remain pivotal areas for ongoing research and development.

Table IX reports the summary of the state of the art works on network monitoring tools.

TABLE IX: Summary on Network Monitoring research works and tools

Network Monitoring and Observation	
Probe traffic generators	Ping [218], Traceroute [219], Apache JMeter [220], Dig [221], iPerf [180], iPerf3 [222], D-ITG [181], Ostinato [182], T-Rex [223], RouteViews [188]
Active Monitoring	Network Delay: [100], [117], [178], [183], [184], [187] Bandwidth: [192]–[194], [224]–[227]
Passive Monitoring	[89], [93], [108], [110], [126], [196], [197], [200], [201]
In-Band Network Telemetry	[114], [216], [217]

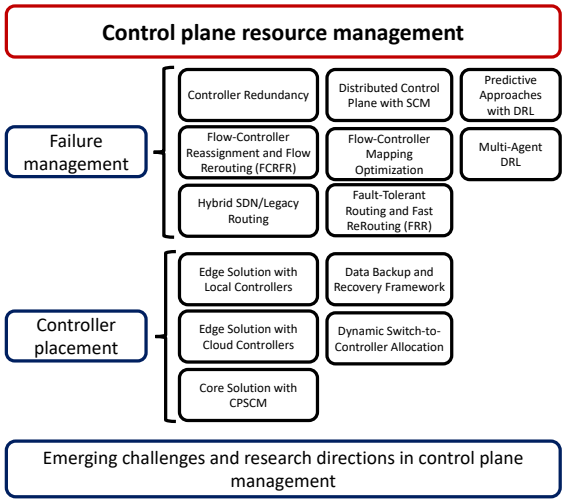

Fig. 12: Taxonomy of control plane resource management.

C. Control plane resource management

Effective coordination between the control and data planes is essential for deploying SD-WAN solutions. The design of CPEs must seamlessly integrate both planes, balancing computational overhead while ensuring scalability and fault isolation. Controllers should be strategically placed to accommodate network growth, minimize propagation delays, and ensure real-time responsiveness. Optimal controller placement is crucial for maintaining network performance and reliability. By addressing these considerations, SD-WAN deployments can achieve enhanced performance, scalability, and reliability, effectively meeting the dynamic demands of modern network environments.

In this section, we delve into various types of failure management and controller placement strategies, providing a comprehensive understanding of these critical features in SD-WAN deployments, as summarized in Fig. 12.

1) Failure management: SD-WAN controller failures pose significant risks to network availability, reliability, and seamless operations. Hardware/software issues, power failures, or unforeseen disruptions can render CPEs and WAN nodes unmanageable, preventing flow adjustments and disrupting network services. Addressing these failures requires redundancy, fault tolerance, and rapid recovery mechanisms.

One widely adopted approach is controller redundancy, which ensures continuity by reassigning switches to backup controllers. The TEACHER framework [228] enhances TE during failures by using a Traffic-aware Path Programmability (TPP) metric for efficient switch-to-controller remapping. However, TEACHER does not guarantee optimal TE in all cases. To enhance TE, Flow-Controller Reassignment and Flow Rerouting (FCRFR) [229] minimizes maximum link utilization by dynamically reallocating flows based on load distribution. A hybrid SDN/legacy approach [230]–[233] improves programmability recovery by combining OpenFlow-based flow tables with Open Short Path First (OSPF) legacy routing. This method allows flows to be rerouted dynamically while maintaining network resilience. Despite its advantages, it increases system complexity and requires real-time state data for precise mapping decisions. Another solution is a distributed control plane [234], where multiple controllers dynamically migrate workloads upon failures thanks to a centralized database that stores flow rules for seamless migration. This ensures smooth failover by detecting failures and reassigning nodes to secondary controllers. However, this approach can introduce significant communication overhead during failover. To mitigate this, fine-grained control mechanisms [235] optimize controller selection by evaluating the number of recovered flows and associated communication costs. Predictive strategies enhance failure resilience by leveraging Deep Reinforcement Learning (DRL) [168], [236]. Techniques such as Deep Q-Networks (DQN) and Deep Deterministic Policy Gradient (DDPG) dynamically redistribute flows among controllers based on network conditions, preventing bottlenecks and optimizing resource utilization. While effective, DRL requires careful tuning of reward functions and extensive training for real-world deployment.

Further improvements come from Multi-Agent DRL (MADRL) [237], where distributed agents (controllers/WAN nodes) collaboratively optimize traffic routing and load balancing. By sharing network state information, MADRL enhances global optimization, proactive decision-making, and dynamic adaptability. However, it introduces challenges such as scalability, synchronization, and coordination overhead. Table X summarizes the different SD-WAN failure management approaches.

2) Controller placement: Controller placement is a critical architectural decision in SD-WAN, directly impacting performance, scalability, and resilience. The approach varies between Edge and Core models, each with distinct benefits and trade-offs.

In Edge solutions, organizations manage their own SD-WAN infrastructure, deploying controllers either on-premises or in the cloud. On-premises controllers minimize latency and enhance security but require significant investment and maintenance efforts [238]. Cloud controllers offer flexibility and scalability but may introduce latency and security concerns due to reliance on external infrastructure.

In Core solutions, controllers are integrated into the provider's WAN infrastructure, ensuring centralized management and optimized resource allocation. Recent research explores fog computing in SD-WAN, using a revenue-oriented service offloading strategy to dynamically allocate resources between fog nodes and the cloud, reducing latency for delay-sensitive applications [239]. A key optimization methodology for Core solutions is the Controller Placement and Switch-Controller Mapping (CPSCM) problem, which minimizes control latency and optimally assigns switches to controllers [140]. CPSCM significantly enhances programmability and decision-making in large-scale deployments but is computationally intensive (NP-hard). Advanced algorithms are required to adapt to dynamic network conditions, such as traffic fluctuations and link failures, adding complexity to real-time operations.

Disaster recovery is fundamental for maintaining network resilience during controller failures. The Data Backup and Recovery (DDBR) method [240] uses an online-offline backup strategy, distributing encrypted controller data across nodes and restoring it via Shamir's secret sharing. While effective in ensuring availability and integrity, it demands high processing and storage resources. Dynamic switch-to-controller allocation further improves SD-WAN by reducing latency and balancing load through real-time congestion awareness [241], though at the cost of significant computational overhead. A robust SD-WAN strategy combines smart controller placement, efficient recovery, and adaptive switch mapping. Key challenges remain in optimizing computation, managing recovery complexity, and scaling allocation in diverse networks. Future work should explore hybrid methods that leverage heuristics, predictive models, and distributed architectures.

3) Emerging challenges and research directions in control plane management: In traditional SD-WAN architectures, a central controller oversees the network's operations by collecting network state information, calculating optimal forwarding paths, and distributing policies to edge devices. While this centralized approach enables fine-grained traffic management, it also presents significant vulnerabilities, including single points of failure, limited scalability, and degraded performance in scenarios involving high latency or network partitioning.

Emerging research proposes a paradigm shift by distributing control plane operations directly to CPEs. This approach eliminates the need for a centralized controller by embedding decision-making capabilities into distributed SD-WAN gateways [242], [243]. In the proposed distributed SD-WAN control plane architecture, each CPE operates autonomously by gathering network statistics, assessing real-time conditions, and executing routing decisions. This approach transforms global policies into localized "inter-site" policies, enabling each CPE to manage data traffic independently. By decentralizing traffic management, the architecture ensures that routing decisions are made locally between connected site pairs, enhancing responsiveness and reducing reliance on a central controller.

The distributed model provides several critical benefits:

TABLE X: Failure management approaches

Methodology	Advantages	Drawbacks	Challenges
Controller Redundancy	Ensures backup controllers are available to take over when primary controllers fail, minimizing downtime.	May offer coarse-grained switch-controller mappings, limiting the recovery scope for inactive flows.	Configuring efficient switch-to-controller mappings and handling resource constraints.
Flow-Controller Reassignment and Flow Rerouting (FCRFR)	Jointly optimizes flow reassignment and rerouting for predictable TE performance.	Computational complexity due to NP-hard optimization problem.	Adapting to dynamic traffic mappings patterns in large-scale networks.
Hybrid SDN/Legacy Routing	Combines OpenFlow and OSPF routing for fine-grained flow programmability recovery.	Adds complexity to routing configuration and relies on commercial SDN switches.	Balancing the allocation of flows between SDN and legacy routing modes while minimizing control overhead.
Distributed Control Plane with SCM	Provides scalability, reliability, and seamless migration of controller workloads through secondary controllers.	Recovery is coarse-grained, limited to switches managed by the failed controller.	Implementing efficient communication between controllers and SCM with minimal latency.
Flow-Controller Mapping Optimization	Ensures fine-grained recovery by reconfiguring control resources for offline flows, optimizing programmability.	Increases communication overhead for real-time decisions.	Selecting optimal controllers for offline flows while minimizing overhead and latency.
Fault-Tolerant Routing and Fast ReRouting (FRR)	Improves resilience by ensuring alternate routes are available during failures.	Adapting traditional IP/MPLS FRR mechanisms to SD-WAN can be complex.	Developing efficient flow-based mathematical models to optimize routing and minimize link utilization.
Predictive Approaches with DRL	Proactively optimizes WAN node processing loads and improves load balancing between controllers.	Requires substantial computational resources and robust training data for the DRL models.	Managing the trade-off between processing complexity and real-time responsiveness in load balancing.
Multi-Agent DRL	Dynamically identifies load conditions and redirects flows for improved load balancing and efficiency.	Higher computational complexity compared to single-agent DRL.	Synchronizing multiple agents and ensuring scalability across large, distributed networks.

TABLE XI: SD-WAN controller placement approaches

Methodology	Advantages	Disadvantages	Challenges
Edge Solution with Local Controllers	- Low latency due to proximity to endpoints. - Enhanced control and security.	- Requires significant upfront investment. - High operational and maintenance costs.	- Scaling infrastructure for distributed networks. - Balancing performance with operational costs.
Edge Solution with Cloud Controllers	- High scalability and flexibility. - Simplified deployment and management by cloud providers.	- Potential latency due to geographical distance. - Security concerns with public cloud reliance.	- Addressing latency for time-sensitive applications. - Mitigating security risks for sensitive data.
Core Solution with CPSCM	- Optimized controller placement reduces latency. - Enhanced programmability and resource allocation.	- Computationally intensive (NP-hard). - Placement strategies may not adapt easily to dynamic conditions.	- Developing scalable algorithms for real-time operations. - Managing dynamic traffic and link failures.
Data Backup and Recovery Framework	- Ensures rapid recovery in case of controller failure. - Preserves data confidentiality and integrity.	- Resource-intensive due to segmentation, encryption, and storage needs. - Operational overhead.	- Ensuring compatibility across heterogeneous network nodes. - Managing resource allocation efficiently.
Dynamic Switch-to-Controller Allocation	- Balances load across controllers. - Reduces latency by adapting to real-time conditions.	- Requires high computational overhead. - Relies heavily on real-time data accuracy.	- Scaling algorithms for large, distributed networks. - Adapting to rapid traffic pattern changes.

- Resilience: By decentralizing control operations, the failure of a single gateway or site does not compromise the entire network. Local decision-making ensures continuous operation and policy enforcement, even under network partitioning.
- Scalability: The elimination of a centralized controller allows the architecture to support thousands of gateways and connections. Each gateway handles only localized decisions, reducing the computational and bandwidth overhead associated with global state synchronization.
- Improved Performance: Distributed control minimizes latency, as decisions are made close to the data sources. This ensures faster response times to changing network conditions, such as link failures or traffic congestion.

However, fully distributed control planes also introduce challenges:

- Consistency Issues: As decisions are made locally, ensuring network-wide policy coherence can be challenging. Mechanisms that prioritize availability over immediate synchronization may lead to eventual consistency, where transient inconsistencies can occur during network state changes until the system converges to a consistent state.
- Increased Complexity: The architecture requires sophisticated algorithms to manage inter-site policy leader roles, resolve conflicts, and distribute decisions efficiently. This complexity may increase operational overhead.
- Communication Overhead: The need for gateways to exchange statistics and decisions introduces additional bandwidth requirements. Although optimization strategies, such as selective sending and reduced frequency updates, mitigate this overhead, it remains a consideration for low-bandwidth environments.

One significant challenge in SD-WAN control plane management is ensuring the dynamic allocation of computational

resources across several CPEs. Advanced selection mechanisms have been proposed, incorporating real-time resource monitoring and weighted scoring systems [244].

Future research should explore hybrid control plane architectures that combine distributed and centralized models to leverage the strengths of both approaches. For instance, a hierarchical design where regional controllers manage clusters of CPEs while delegating localized decision-making to individual CPEs could balance resilience and efficiency. Additionally, the integration of AI-driven decision-making frameworks can enhance the adaptability of distributed gateways, enabling predictive traffic management and proactive fault mitigation. By focusing on these directions, the distributed control plane model for SD-WAN can evolve into a robust and scalable solution capable of meeting the demands of modern, dynamic networking environments.

VII. Lessons Learned

This survey has explored the multifaceted landscape of SD-WAN, highlighting its transformative impact on modern networking. By analyzing its core functionalities, real-world applications, and ongoing research, several key lessons have emerged.

Cost Efficiency and Simplification. SD-WAN significantly reduces operational costs by replacing traditional MPLS-VPNs with cost-effective broadband links. Its centralized control framework simplifies network management, optimizing resource utilization and minimizing downtime.

Scalability and Flexibility. SD-WAN seamlessly adapts to dynamic traffic demands and geographically distributed infrastructures, ensuring scalable deployments across industries. Its integration with edge computing and cloud services further enhances its flexibility.

Enhanced QoS and Resilience. Intelligent path selection and dynamic load balancing strengthen network resilience, ensuring consistent QoS for latency-sensitive applications such as real-time analytics and video conferencing. Its fault-tolerant architecture mitigates link failures and optimizes performance.

Traffic Engineering Approaches. TE in SD-WAN varies across heuristic and ML-based models. RL-driven TE offers high adaptability in dynamic environments but requires extensive training data and computational resources. In contrast, heuristic-based TE is computationally efficient and ideal for predictable traffic patterns but lacks adaptability in rapidly changing networks.

Challenges in Multi-Vendor Integration. Interoperability remains a significant hurdle in multi-vendor SD-WAN environments. The lack of standardized frameworks and APIs leads to inconsistent implementations, complicating seamless integration across platforms.

Advanced Monitoring Insights. AI-driven monitoring frameworks and INT-based telemetry enhance real-time fault detection and proactive maintenance. However, they introduce computational complexity, necessitating optimization strategies for scalability and efficiency.

These insights shape SD-WAN's trajectory, influencing its integration with hybrid cloud environments and decentralized architectures. Future research focuses on its intersection with 6G, AI-driven optimization, and energy-efficient designs, addressing challenges such as ultra-low latency and enhanced security. By consolidating these lessons, SD-WAN continues to evolve as a robust, adaptive, and secure networking solution across diverse application domains.

VIII. Final Remarks and Future Perspective

The future of SD-WAN is poised to intersect with cutting-edge technologies, including Satellite 6G, the Metaverse, Free Space Optics (FSO), and Networked Music Performance (NMP). These advancements will redefine connectivity, application performance, and network intelligence. In the era of 6G, networks will deliver unprecedented speed, ultra-low latency, and seamless integration with edge computing and network slicing [245]. SD-WAN will play a pivotal role by enabling intelligent traffic orchestration, dynamic path selection, and application-aware network slicing [246]. Furthermore, its integration with Satellite 6G will enhance global connectivity, addressing challenges in remote and underserved areas [247]. SD-WAN's adaptability extends to the Metaverse, where immersive digital environments demand scalable, high-performance networking [248]. Additionally, FSO technology presents a promising alternative to traditional wired networks, particularly in urban and remote deployments [249]. In applications such as NMP, SD-WAN's real-time resource optimization capabilities can minimize latency and enhance QoE for distributed artistic collaborations [250]. Future advancements in SD-WAN will focus on network automation, leveraging AI-driven algorithms and zero-touch provisioning to reduce human intervention and enhance scalability. A notable development is the integration of generative AI. Gartner predicts that by 2027, 70% of network operations personnel will rely on AI for SD-WAN automation, up from less than 5% in early 2024 [251]. Leading vendors are embedding AI into SD-WAN solutions, offering AI-assisted configuration management, incident resolution, and real-time analytics to improve operational efficiency. Despite its potential, SD-WAN faces challenges such as integration complexity with existing networks, managing security in distributed architectures, and ensuring scalability to accommodate growing data traffic and evolving network services.

A visionary future for SD-WAN. Beyond incremental advancements, SD-WAN could extend beyond Earth, orchestrating interplanetary communication networks for future Mars colonization and lunar habitats [252], [253]. Its adaptability could support a "universal connectivity fabric," enabling real-time data transmission between planetary networks using advanced quantum networking. At a more granular level, SD-WAN may evolve into bio-integrated networks, leveraging nanotechnology for seamless human-to-machine communication [254]. Embedded personal devices at the cellular level could dynamically interface with global networks, with SD-WAN prioritizing life-critical traffic in real time. Further, into the future, SD-WAN could become a foundational element of digital consciousness, facilitating real-time neural networking within the Metaverse [255]. Thought itself could become a

form of data, with SD-WAN ensuring seamless connectivity across a globally interlinked society. In conclusion, SD-WAN stands at the forefront of a networking revolution, bridging present and future technological paradigms. Its synergy with emerging technologies will shape a hyper-connected world where intelligent, adaptive, and secure networking will be paramount. As SD-WAN continues to evolve, its role in enabling next-generation digital applications, from terrestrial to interplanetary communications, will solidify its place as an indispensable asset in the evolution of global networks.

ACKNOWLEDGMENT

This work was supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP D43C22003080001, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").

REFERENCES

[1] E. Banks, "Software-defined wan: A primer," 2014, Accessed: 2025-01-09. [Online]. Available: https://www.networkcomputing.com/networking/software-defined-wan-primer

[2] "ATM technical specifications," ATM Forum, 1999.

[3] P. Smith, *Frame relay: principles and applications*. Addison-Wesley Longman Publishing Co., Inc., 1993.

[4] B. S. Davie and Y. Rekhter, *MPLS: technology and applications*. Morgan Kaufmann Publishers Inc., 2000.

[5] T. Benson, A. Akella, and D. Maltz, "Unraveling the complexity of network management," in *Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation*, ser. NSDI'09. USA: USENIX Association, 2009, p. 335–348.

[6] R. King, "Gap connects stores over the internet with software-defined networking," 2015, Accessed: 2025-01-09. [Online]. Available: https://www.wsj.com/articles/BL-CIOB-8431

[7] "Mclaren and ntt com to deploy new sdx technology at the formula 1 japanese grand prix 2018," NTT, 2018, Accessed: 2025-01-09. [Online]. Available: https://www.ntt.com/en/about-us/press-releases/news/article/2018/1004.html

[8] M. Kamaludeen, S. Ismaeel, F. Petrocelli, C. Carrington, and C. Scarfo, "Data-driven technique to estimate the required broadband speed for k-12 schools," in *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*. IEEE, 2017, pp. 99–104.

[9] I.-A. Ivanciu, R. Botez, C.-M. Iurian, A.-V. Dumitrescu, T.-M. Blaga, and V. Dobrota, "An sd-wan approach for eut+ network," in *2021 20th RoEduNet Conference: Networking in Education and Research (RoEduNet)*. IEEE, 2021, pp. 1–6.

[10] M. Savi, F. Sartori, and R. Melen, "Rethinking the design of wearable expert systems: the role of network infrastructures," in *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2020, pp. 392–395.

[11] A. N. Fitriani, T. Raharjo, B. Hardian, and A. Prasetyo, "It infrastructure agile adoption for sd-wan project implementation in pharmaceutical industry: Case study of an indonesian company," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2021, pp. 1–6.

[12] W. Pratiwi and D. Gunawan, "Design and strategy deployment of sd-wan technology: in indonesia (case study: Pt. xyz)," in *2021 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*. IEEE, 2021, pp. 1–6.

[13] R. Asif and K. Ghanem, "Ai secured sd-wan architecture as a latency critical iot enabler for 5g and beyond communications," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2021, pp. 1–6.

[14] Y. Ushakov, M. Ushakova, and L. Legashev, "Problems of building infrastructure vehicular ad hoc networks based on sd-wan technologies," in *2022 International Siberian Conference on Control and Communications (SIBCON)*. IEEE, 2022, pp. 1–4.

[15] G. Maier, A. Albanese, M. Ciavotta, N. Ciulli, S. Giordano, E. Giusti, A. Salvatore, and G. Schembra, "Watchedge: Smart networking for distributed ai-based environmental control," *Computer Networks*, p. 110248, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S138912862400080X

[16] A. Lara, A. Kolasani, and B. Ramamurthy, "Network innovation using openflow: A survey," *IEEE communications surveys & tutorials*, vol. 16, no. 1, pp. 493–512, 2013.

[17] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2014.

[18] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and openflow: From concept to implementation," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2181–2206, 2014.

[19] R. Alvizu, G. Maier, N. Kukreja, A. Pattavina, R. Morro, A. Capello, and C. Cavazzoni, "Comprehensive survey on t-sdn: Software-defined networking for transport networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2232–2283, 2017.

[20] A. Blenk, A. Basta, M. Reisslein, and W. Kellerer, "Survey on network virtualization hypervisors for software defined networking," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 655–685, 2015.

[21] M. Shirazipour, Y. Zhang, N. Beheshti, G. Lefebvre, and M. Tatipamula, "Openflow and multi-layer extensions: Overview and next steps," in *2012 European Workshop on Software Defined Networking*. IEEE, 2012, pp. 13–17.

[22] J.-P. Elbers and A. Autenrieth, "From static to software-defined optical networks," in *2012 16th International Conference on Optical Network Design and Modelling (ONDM)*. IEEE, 2012, pp. 1–4.

[23] R. Alvizu and G. Maier, "Can open flow make transport networks smarter and dynamic," in *An overview on transport SDN. Paper presented at: 2014 International Conference on Smart Communications in Network Technologies (SaCoNeT)*, 2014.

[24] J. R. de Almeida Amazonas, G. Santos-Boada, and J. Solé-Pareta, "A critical review of openflow/sdn-based networks," in *2014 16th International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2014, pp. 1–5.

[25] C. S. Kankipati, K. Kancharla, N. S. Rampalli, S. Bandi, and R. R. Chintala, "The role of sd-wan in cloud connectivity and digital transformation," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 2023, pp. 1169–1173.

[26] F. Bannour, S. Souihi, and A. Mellouk, "Distributed sdn control: Survey, taxonomy, and challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 333–354, 2018.

[27] C. Fu, B. Wang, and W. Wang, "Software-defined wide area networks (sd-wans): A survey," *Electronics*, vol. 13, no. 15, p. 3011, 2024.

[28] O. Michel and E. Keller, "Sdn in wide-area networks: A survey," in *2017 Fourth International Conference on Software Defined Systems (SDS)*, 2017, pp. 37–42.

[29] K. G. Yalda, D. J. Hamad, and N. Ţăpuş, "A survey on software-defined wide area network (sd- wan) architectures," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–5.

[30] M. A. Ouamri, T. Alharbi, D. Singh, and Z. Sylia, "A comprehensive survey on software-defined wide area network (sd-wan): principles, opportunities and future challenges," *The Journal of Supercomputing*, vol. 81, no. 1, pp. 1–47, 2025.

[31] "SD-WAN Service Components Graphic," MEF, Accessed: 2025-01-09. [Online]. Available: https://wiki.mef.net/display/CESG/SD-WAN+Service+Components+Graphic

[32] R. C. Palancar, R. A. L. da Silva, J. L. F. Chavarría, D. R. López, A. J. E. Armengol, and R. G. Tinoco, "Virtualization of residential customer premise equipment. lessons learned in brazil vcpe trial," *it - Information Technology*, vol. 57, no. 5, pp. 285–294, 2015. [Online]. Available: https://doi.org/10.1515/itit-2015-0028

[33] T.-H. Nguyen, T. Nguyen, and M. Yoo, "Analysis of deployment approaches for virtual customer premises equipment," in *2018 International Conference on Information Networking (ICOIN)*, 2018, pp. 289–291.

[34] "NFV, Network Functions Virtualisation, and Use Cases," ETSI GS NFV, 2013.

[35] J. Proença, T. Cruz, P. Simões, M. Freitas, and R. Calé, "Virtualizing customer premises equipment," in *2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2021, pp. 104–105.

[36] A. Kimura, S. Kawano, H. Tsuchiya, S. Homma, and A. Okada, "Evaluation of virtual customer premises equipment prototype system with open source software," in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, 2018, pp. 1–3.

[37] "Health insurance portability and accountability act of 1996," U.S. Department of Health and Human Services, 1996, Accessed: 2025-01-09. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[38] A. Menko, "2023 broadband pricing index," US Telecom, 2023, Accessed: 2025-01-09. [Online]. Available: https://ustelecom.org/wp-content/uploads/2023/10/USTelecom-2023-BPI-Report-final.pdf

[39] D. Howdle, "Global broadband pricing league table 2024," Cable, Accessed: 2025-01-09. [Online]. Available: https://www.cable.co.uk/broadband/pricing/worldwide-comparison/

[40] "Gps fleet tracking system," HOS247, Accessed: 2025-01-09. [Online]. Available: https://hos247.com/gps-fleet-tracking/

[41] Force Fleet Tracking, Accessed: 2025-01-09. [Online]. Available: https://forcefleettracking.com/

[42] V. Cheimaras, N. Peladarinos, N. Monios, S. Daousis, S. Papagiakoumos, P. Papageorgas, and D. Piromalis, "Emergency communication system based on wireless lpwan and sd-wan technologies: A hybrid approach," *Signals*, vol. 4, no. 2, pp. 315–336, 2023. [Online]. Available: https://www.mdpi.com/2624-6120/4/2/17

[43] J. Conde, G. Martínez, P. Reviriego, and J. A. Hernández, "Round trip times (rtts): Comparing terrestrial and leo satellite networks," in *2024 27th Conference on Innovation in Clouds, Internet and Networks (ICIN)*. IEEE, 2024, pp. 42–46.

[44] "Versa SD-WAN Solution Use Case for Satellite ISPs," Versa Networks, Accessed: 2025-01-09. [Online]. Available: https://versa-networks.com/documents/white-papers/versa-sd-wan-solution-use-case-for-satellite-isps.pdf

[45] "Establishing remote connection for mining facilities in west africa," NeXat, Accessed: 2025-01-09. [Online]. Available: https://www.nexat.be/case-studies/remote-connection-for-mining-facilities-in-west-africa

[46] "Viasat enterprise," Viasat, Accessed: 2025-01-09. [Online]. Available: https://www.viasat.com/enterprise/

[47] S. Ma, Y. C. Chou, H. Zhao, L. Chen, X. Ma, and J. Liu, "Network characteristics of leo satellite constellations: A starlink-based measurement from end users," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.

[48] A. M. Frank Rayal, Riad Hartani, "Defining the synergies between leo satellite constellations and submarine cables," Xona Partners, 2020, Accessed: 2025-01-09. [Online]. Available: https://xonapartners.com/wp-content/uploads/2020/12/Defining-the-Synergies-between-LEO-Satellite-Constellations-and-Submarine-Cables.pdf

[49] "Starlink specifications," Starlink, accessed: 2025-01-09. [Online]. Available: https://www.starlink.com/legal/documents/DOC-1400-28829-70

[50] "OneWeb's LEO Satellites Clock 400 Mbit/s, Latency of 32 MS in Initial Tests," LightReading, 2019, accessed: 2025-01-09. [Online]. Available: https://www.lightreading.com/satellite/oneweb-s-leo-satellites-clock-400-mbit-s-latency-of-32-ms-in-initial-tests

[51] G. Sguotti, S. Troia, and G. Maier, "Mitigating satellite link variability in sd-wan using constellation-aware edge routing," in *2025 25th International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2025.

[52] "Multi-Orbit SD-WAN for Hybrid Networks," XipLink, Accessed: 2025-01-09. [Online]. Available: https://www.xiplink.com/solutions/multi-orbit-sd-wan

[53] "New SD-WAN connectivity solution from neXat simplifies network management and enhances critical communications," NeXat, Accessed: 2025-01-09. [Online]. Available: https://www.nexat.be/news/new-sd-wan-solution

[54] "Subscriber Layer 1 Service Attributes Technical Specification," MEF Forum, 2018, Accessed: 2025-01-09. [Online]. Available: https://www.mef.net/wp-content/uploads/2018/08/MEF-63.pdf

[55] "IP Service Attributes," MEF Forum, 2019.

[56] "Subscriber Ethernet Services Definitions," MEF Forum, 2019.

[57] M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, C. Wright, "Virtual extensible local area network (vxlan): A framework for overlaying virtualized layer 2 networks over layer 3 networks," Internet Requests for Comments, 08 2014. [Online]. Available: https://www.hjp.at/doc/rfc/rfc7348.html

[58] W. Townsley, A. Valencia, A. Rubens, G. Pall, G. Zorn, B. Palter, "Layer two tunneling protocol "l2tp"," Internet Requests for Comments, RFC Editor, RFC 2661, 08 1999. [Online]. Available: https://tools.ietf.org/html/rfc2661

[59] J. Lau, M. Townsley, I. Goyret, "Layer two tunneling protocol - version 3 (l2tpv3)," Internet Requests for Comments, RFC Editor, RFC 3931, 03 2005. [Online]. Available: https://tools.ietf.org/html/rfc3931

[60] W. Simpson, "The point-to-point protocol (ppp)," Internet Requests for Comments, RFC Editor, RFC 1661, 07 1994. [Online]. Available: https://tools.ietf.org/html/rfc1661

[61] S. Hanks, T. Li, D. Farinacci, P. Traina, "Generic routing encapsulation (gre)," Internet Requests for Comments, RFC Editor, RFC 1701, 10 1994. [Online]. Available: https://www.rfc-editor.org/rfc/rfc1701.html

[62] S. Kent, R. Atkinson, "Ip encapsulating security payload (esp)," Internet Requests for Comments, RFC Editor, RFC 2406, 11 1998. [Online]. Available: https://tools.ietf.org/html/rfc2406#section-1

[63] W. Stallings, *Cryptography and Network Security*. Pearson Education India, 2006.

[64] J. Wang, M. Bewong, and L. Zheng, "Sd-wan: Hybrid edge cloud network between multi-site sddc," *Computer Networks*, vol. 250, p. 110509, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128624003414

[65] R. Alvizu, G. Maier, N. Kukreja, A. Pattavina, R. Morro, A. Capello, and C. Cavazzoni, "Comprehensive survey on t-sdn: Software-defined networking for transport networks," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2232–2283, 2017.

[66] G. Mine, J. Hai, L. Jin, and Z. Huiying, "A design of sd-wan-oriented wide area network access," in *2020 International Conference on Computer Communication and Network Security (CCNS)*. IEEE, 2020, pp. 174–177.

[67] Ryu Controller, Accessed: 2025-01-09. [Online]. Available: https://ryu-sdn.org/

[68] P. Berde, M. Gerola, J. Hart, Y. Higuchi, M. Kobayashi, T. Koide, B. Lantz, B. O'Connor, P. Radoslavov, W. Snow *et al.*, "Onos: towards an open, distributed sdn os," in *Proceedings of the third workshop on Hot topics in software defined networking*, 2014, pp. 1–6.

[69] J. Medved, R. Varga, A. Tkacik, and K. Gray, "Opendaylight: Towards a model-driven sdn controller architecture," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, 2014, pp. 1–6.

[70] Floodlight SDN controller, Accessed: 2025-01-09. [Online]. Available: https://github.com/floodlight/floodlight

[71] R. Vilalta, R. Muñoz, R. Casellas, R. Martínez, V. López, O. G. de Dios, A. Pastor, G. P. Katsikas, F. Klaedtke, P. Monti, A. Mozo, T. Zinner, H. Øverby, S. Gonzalez-Diaz, H. Lønsethagen, J.-M. Pulido, and D. King, "Teraflow: Secured autonomic traffic management for a tera of sdn flows," in *2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)*, 2021, pp. 377–382.

[72] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM computer communication review*, vol. 38, no. 2, pp. 69–74, 2008.

[73] M. Bjorklund, "Yang-a data modeling language for the network configuration protocol (netconf)," Tech. Rep., 2010.

[74] A. Bierman, M. Bjorklund, and K. Watsen, "Restconf protocol," Tech. Rep., 2017.

[75] Open Source NFV Management and Orchestration (MANO), Accessed: 2025-01-09. [Online]. Available: https://osm.etsi.org/

[76] Open Network Automation Platform (ONAP), Accessed: 2025-01-09. [Online]. Available: https://www.onap.org/

[77] Cloudify, Accessed: 2025-01-09. [Online]. Available: https://cloudify.co/

[78] R. Perez, A. Zabala, and A. Banchs, "Alviu: An intent-based sd-wan orchestrator of network slices for enterprise networks," in *2021 IEEE 7th International Conference on Network Softwarization (NetSoft)*. IEEE, 2021, pp. 211–215.

[79] Z. Yang, Y. Cui, B. Li, Y. Liu, and Y. Xu, "Software-defined wide area network (sd-wan): Architecture, advances and opportunities," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, 2019, pp. 1–9.

[80] P. Segeč, M. Moravčik, J. Uratmová, J. Papán, and O. Yeremenko, "Sd-wan - architecture, functions and benefits," in *2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2020, pp. 593–599.

[81] A. Lara, B. Ramamurthy, E. Pouyoul, and I. Monga, "Wan virtualization and dynamic end-to-end bandwidth provisioning using sdn," in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2015, pp. 1–3.

[82] J. Carapinha, J. Bonnet, B. Parreira, J. Silva, I. Trajkovska, T. M. Bohnert, G. Xilouris, M.-A. Kourtis, and C. Sakkas, "Deployment of virtual network functions over multiple wan interconnected pops,"

in *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. IEEE, 2017, pp. 252–257.

[83] L. M. Contreras, L. Luque, G. Landi, G. Bernini, G. Carrozzo, J. García-Reinoso, and M. M. Roselló, "Interworking of softwarized infrastructures for enabling 5g multi-site slice orchestration," in *2019 IEEE Conference on Network Softwarization (NetSoft)*. IEEE, 2019, pp. 458–463.

[84] S. Simpson, A. Farshad, P. McCherry, A. Magzoub, W. Fantom, C. Rotsos, N. Race, and D. Hutchison, "Dataplane broker: Open wan control for multi-site service orchestration," in *2019 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. IEEE, 2019, pp. 1–6.

[85] P. Karamichailidis, K. Choumas, and T. Korakis, "Enabling multi-domain orchestration using open source mano, openstack and opendaylight," in *2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. IEEE, 2019, pp. 1–6.

[86] A. Bravalheri, A. S. Muqaddas, N. Uniyal, R. Casellas, R. Nejabati, and D. Simeonidou, "Vnf chaining across multi-pops in osm using transport api," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2019, pp. 1–3.

[87] "Magic quadrant for wan edge infrastructure," Gartner, 2021, Accessed: 2025-01-09. [Online]. Available: https://www.gartner.com/en/documents/4005922

[88] "Managed sd-wan solutions," AT&T, Accessed: 2025-01-09. [Online]. Available: https://www.business.att.com/products/sd-wan.html

[89] K. Golani, K. Goswami, K. Bhatt, and Y. Park, "Fault tolerant traffic engineering in software-defined wan," pp. 01 205–01 210, 2018.

[90] "Northstar controller," Juniper Networks, Accessed: 2025-01-09. [Online]. Available: https://www.juniper.net/us/en/products/network-automation/northstar-controller.html

[91] "Juniper northstar api documentation," Juniper Networks, Accessed: 2025-01-09. [Online]. Available: http://www.juniper.net/techpubs/en_US/northstar2.1.0/informationproducts/api-ref/api-ref.html

[92] M. Shojaee, M. Neves, and I. Haque, "Safeguard: Congestion and memory-aware failure recovery in sd-wan," in *2020 16th International Conference on Network and Service Management (CNSM)*. IEEE, 2020, pp. 1–7.

[93] L. Vdovin, P. Likin, and A. Vilchinskii, "Network utilization optimizer for sd-wan," pp. 1–4, 2014.

[94] Y. Chen, Q. Wu, W. Zhang, and Q. Liu, "Sd-wan source route based on protocol-oblivious forwarding," pp. 95–99, 2018.

[95] C. yao Hong, S. Mandal, M. A. Alfares, M. Zhu, R. Alimi, K. N. Bollineni, C. Bhagat, S. Jain, J. Kaimal, J. Liang, K. Mendelev, S. Padgett, F. T. Rabe, S. Ray, M. Tewari, M. Tierney, M. Zahn, J. Zolla, J. Ong, and A. Vahdat, "B4 and after: Managing hierarchy, partitioning, and asymmetry for availability and scale in google's software-defined wan," in *SIGCOMM'18*, 2018. [Online]. Available: https://conferences.sigcomm.org/sigcomm/2018/program_tuesday.html

[96] E. Biederman, "Mininet," Accessed: 2025-01-09. [Online]. Available: https://mininet.org/

[97] S. Xu, M. Kodialam, T. Lakshman, and S. S. Panwar, "Tomography based learning for load distribution through opaque networks," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 656–670, 2021.

[98] L. Liu, L. Chen, H. Xu, and H. Shao, "Automated traffic engineering in sdwan: Beyond reinforcement learning," pp. 430–435, 2020.

[99] O. Lemeshko, M. Yevdokymenko, O. Yeremenko, A. M. Hailan, P. Segeč, and J. Papán, "Design of the fast reroute qos protection scheme for bandwidth and probability of packet loss in software-defined wan," pp. 1–5, 2019.

[100] D. Z. Tootaghaj, F. Ahmed, P. Sharma, and M. Yannakakis, "Homa: An efficient topology and route management approach in sd-wan overlays," pp. 2351–2360, 2020.

[101] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring isp topologies with rocketfuel," *IEEE/ACM Transactions on networking*, vol. 12, no. 1, pp. 2–16, 2004.

[102] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1765–1775, 2011.

[103] W. Cheng, F. Yang, T. Han, R. Han, W. Jiang, G. Zhang, J. Yang, and X. Zhang, "The smart sd-wan architecture," in *2022 IEEE 22nd International Conference on Communication Technology (ICCT)*. IEEE, 2022, pp. 667–671.

[104] L. Song, Y. Jin, P. Wang, D. Ma, W. Chen, and L. Cui, "Multi-path routing deployment method based on srv6," in *2021 IEEE Intl Conf on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking (ISPA/BDCloud/SocialCom/SustainCom)*, 2021, pp. 723–730.

[105] 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. 1067–1081, 2024.

[106] C. Scarpitta, P. L. Ventre, F. Lombardo, S. Salsano, and N. Blefari-Melazzi, "Everywan-an open source sd-wan solution," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. IEEE, 2021, pp. 1–7.

[107] "Aws cloud wan parries google, microsoft," SDxCentral, 2021, Accessed: 2025-01-09. [Online]. Available: https://www.sdxcentral.com/articles/news/aws-cloud-wan-parries-google-microsoft/2021/12/

[108] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu *et al.*, "B4: Experience with a globally-deployed software defined wan," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 3–14, 2013.

[109] C.-Y. Hong, S. Kandula, R. Mahajan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer, "Achieving high utilization with software-driven wan," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, 2013, pp. 15–26.

[110] K. Phemius and M. Bouet, "Implementing openflow-based resilient network services," pp. 212–214, 2012.

[111] R. E. Mora-Huiracocha, P. L. Gallegos-Segovia, P. E. Vintimilla-Tapia, J. F. Bravo-Torres, E. J. Cedillo-Elias, and V. M. Larios-Rosillo, "Implementation of a sd-wan for the interconnection of two software defined data centers," pp. 1–6, 2019.

[112] V.-V. Coman and V. Dobrota, "Open-source software-defined wide area network: An inter-cloud approach," in *2022 International Symposium on Electronics and Telecommunications (ISETC)*. IEEE, 2022, pp. 1–4.

[113] D. Farinacci, T. Li, S. Hanks, D. Meyer, and P. Traina, "Generic routing encapsulation (gre)," Tech. Rep., 2000.

[114] Y.Zhang, T. Pan, and et al., "Vxlan-based int: In-band network telemetry for overlay network monitoring," *IEEE INFOCOM 2021*, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9484508

[115] A. A. Ansari, S. P. Rachuri, A. A. Kherani, and D. Tandur, "An sd-wan controller for delay jitter minimization in coded multi-access systems," pp. 1–6, 2019.

[116] N. Nikaein, R. Knopp, F. Kaltenberger, L. Gauthier, C. Bonnet, D. Nussbaum, and R. Ghaddab, "Openairinterface: an open lte network in a pc," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 305–308.

[117] S. Troia, L. M. M. Zorello, A. J. Maralit, and G. Maier, "Sd-wan: an open-source implementation for enterprise networking services," in *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2020, pp. 1–4.

[118] G. Malkin, "Rip version 2," Tech. Rep., 1998.

[119] B. Pfaff, J. Pettit, T. Koponen, E. Jackson, A. Zhou, J. Rajahalme, J. Gross, A. Wang, J. Stringer, P. Shelar *et al.*, "The design and implementation of open {vSwitch}," in *12th USENIX symposium on networked systems design and implementation (NSDI 15)*, 2015, pp. 117–130.

[120] A. Botta, A. Dainotti, and A. Pescapé, "A tool for the generation of realistic network workload for emerging networking scenarios," *Computer Networks*, vol. 56, no. 15, pp. 3531–3547, 2012.

[121] G. Zakurdaev, M. Ismail, and C.-H. Lung, "Dynamic on-demand virtual extensible lan tunnels via software-defined wide area networks," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0995–1000.

[122] H. H. G. Guardabaxo and G. S. Pavani, "Qoe management of http/2 traffic in software-defined wide area networks," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2022, pp. 1–5.

[123] S. Lee, K.-Y. Chan, and T.-Y. Chen, "Design and implementation of an sd-wan vpn system to support multipath and multi-wan-hop routing in the public internet," 2020.

[124] A. Botta, R. Canonico, A. Navarro, S. Ruggiero, and G. Ventre, "Ai-enabled sd-wan: the case of reinforcement learning," in *2022 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2022, pp. 1–6.

[125] R. Y. Manova, E. Sukmadirana, and N. S. Nurmanah, "Comparative analysis of quality of service and performance of mpls, eoip and sd-wan," in *2022 1st International Conference on Information System & Information Technology (ICISIT)*. IEEE, 2022, pp. 403–408.

[126] S. Troia, M. Mazzara, M. Savi, L. M. M. Zorello, and G. Maier, "Resilience of delay-sensitive services with transport-layer monitoring in sd-wan," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2652–2663, 2022.

[127] O. Fares, A. Dandoush, and N. Aitsaadi, "Sdn-based platform enabling intelligent routing within transit autonomous system networks," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2022, pp. 909–912.

[128] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2003.

[129] D. Henrici and L. Wischhof, "Site connectivity with towalink–implementing an open source" sd-wan light"," in *2022 30th Telecommunications Forum (TELFOR)*. IEEE, 2022, pp. 1–4.

[130] J. A. Donenfeld, "Wireguard: next generation kernel network tunnel." in *NDSS*, 2017, pp. 1–12.

[131] G. Sguotti, S. Troia, M. Grieco, and G. Maier, "An open source sd-wan cpe with fast packet processing and secure connectivity," in *2025 IEEE 26th International Conference on High Performance Switching and Routing (HPSR)*, 2025.

[132] Y. Zhang, J. Tourrilhes, Z.-L. Zhang, and P. Sharma, "Improving sd-wan resilience: From vertical handoff to wan-aware mptcp," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 347–361, 2021.

[133] Y. Magnouche, P. T. A. Quang, J. Leguay, X. Gong, and F. Zeng, "Distributed load balancing from the edge in ip networks," pp. 1–6, 2021.

[134] Y. Zhang, H. Zhang, P. Cong, W. Wang, and K. Xu, "Grandet: Cost-aware traffic scheduling without prior knowledge in sd-wan," in *2023 IEEE/ACM 31st International Symposium on Quality of Service (IWQoS)*. IEEE, 2023, pp. 1–10.

[135] W. Wang, H. Wang, G. Wu, X. Liang, W. Chen, and Y. Feng, "Research on the application of sd-wan technology in power communication scenarios," in *2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT)*. IEEE, 2022, pp. 720–723.

[136] Z. Duliński, R. Stankiewicz, G. Rzym, and P. Wydrych, "Dynamic traffic management for sd-wan inter-cloud communication," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1335–1351, 2020.

[137] S. Troia, M. Mazzara, L. M. M. Zorello, and A. Pattavina, "Resiliency in sd-wan with ebpf monitoring: municipal network and video streaming use cases," in *2021 17th international conference on the design of reliable communication networks (DRCN)*. IEEE, 2021, pp. 1–3.

[138] B. Koné and A. D. Kora, "Management and orchestration for network function virtualization in a voip testbed: a multi-domain case," in *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2021, pp. 372–376.

[139] P. T. A. Quang, J. Leguay, F. Zeng, J. Hou, B. Yu, and D. Restivo, "Dynamic qos for high quality sd-wan overlays," in *NOMS 2024-2024 IEEE Network Operations and Management Symposium*. IEEE, 2024, pp. 1–3.

[140] S. Dou, L. Qi, C. Yao, and Z. Guo, "Exploring the impact of critical programmability on controller placement for software-defined wide area networks," *IEEE/ACM Transactions on Networking*, 2023.

[141] I. D. Emmanuel, N. Linge, and S. Hill, "Analysis of sd-wan packets using machine learning algorithm," in *2023 Conference on Information Communications Technology and Society (ICTAS)*. IEEE, 2023, pp. 1–6.

[142] M. Finsterbusch, C. Richter, E. Rocha, J.-A. Muller, and K. Hanssgen, "A survey of payload-based traffic classification approaches," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1135–1156, 2014.

[143] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, C. Wang, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (sdn): Research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 393–430, 2018.

[144] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE access*, vol. 6, pp. 32 328–32 338, 2018.

[145] Z. Nazemi Absardi and R. Javidan, "A predictive sd-wan traffic management method for iot networks in multi-datacenters using deep rnn," *IET Communications*, vol. 18, no. 18, pp. 1151–1165, 2024.

[146] J. Li, Z. Xu, J. Wu, J. Yu, and Y. Zhou, "Coordination control approach of data flow for geographically distributed charging stations," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2022, pp. 1–5.

[147] M. Ghaderi, W. Liu, S. Xiao, and F. Li, "Learning traffic encoding matrices for delay-aware traffic engineering in sd-wans," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2022, pp. 1–9.

[148] L. Borgianni, S. Troia, D. Adami, G. Maier, and S. Giordano, "From mpls to sd-wan to ensure qos and qoe in cloud-based applications," in *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*. IEEE, 2023, pp. 366–369.

[149] A. Botta, R. Canonico, A. Navarro, G. Stanco, and G. Ventre, "Scalable reinforcement learning for dynamic overlay selection in sd-wans," in *2023 IFIP Networking Conference (IFIP Networking)*. IEEE, 2023, pp. 1–9.

[150] P. T. A. Quang, J. Leguay, X. Gong, and X. Huiying, "Global qos policy optimization in sd-wan," *arXiv preprint arXiv:2304.05473*, 2023.

[151] Y. Gao, "Line quality evaluation and service scheduling scheme based on sd-wan network," in *2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence (SEAI)*. IEEE, 2023, pp. 240–244.

[152] L. Borgianni, D. Adami, S. Giordano, and M. Pagano, "Enhancing reliability in rural networks using a software-defined wide area network," *Computers*, vol. 13, no. 5, 2024. [Online]. Available: https://www.mdpi.com/2073-431X/13/5/113

[153] P. T. A. Quang, S. Martin, J. Leguay, X. Gong, and X. Huiying, "Intent-based routing policy optimization in sd-wan," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4914–4919.

[154] G. Sguotti, S. Troia, and G. Maier, "On the service-oriented availability analysis of software defined wide area network," in *2025 IEEE 11th International Conference on Network Softwarization (NetSoft)*. IEEE, 2025.

[155] "Calculate network availability with multiple wan circuits," SASE Experts, 2016, Accessed: 2025-01-09. [Online]. Available: https://www.sase-experts.com/blog/calculate-network-availability-multiple-wan-circuits/

[156] D. Greenfield, "Ensuring high uptime with sd-wan," Cato Networks, 2018, Accessed: 2025-01-09. [Online]. Available: https://web.archive.org/web/20230530145318/https://www.catonetworks.com/blog/ensuring-high-uptime-with-sd-wan/

[157] S. Troia, F. Sapienza, L. Varé, and G. Maier, "On deep reinforcement learning for traffic engineering in sd-wan," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2198–2212, 2020.

[158] L. Borgianni, S. Troia, D. Adami, G. Maier, and S. Giordano, "Assessing the efficacy of reinforcement learning in enhancing quality of service in sd-wans," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 1765–1770.

[159] Dijkstra, Edsger W, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[160] M. Ye, B. Han, X. Fang, X. Zou, X. Bao, X. Xie, S. Xiao, Y. Lin, and H. J. Chao, "Dynamic path switching for traffic engineering in sd-wan with ebpf," in *2025 IEEE 26th International Conference on High Performance Switching and Routing (HPSR)*. IEEE, 2025.

[161] A. Hosseini, M. Dolati, and M. Ghaderi, "Bulk transfer scheduling with deadline in best-effort sd-wans," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 313–321.

[162] S. S. Reddy, N. Pitaev, and A. Leivadeas, "Informed network routing for sd-wan enabled saas optimization: An m365 proof of concept," in *2024 7th Conference on Cloud and Internet of Things (CIoT)*. IEEE, 2024, pp. 1–7.

[163] Y. Xin and Y. Wang, "Partitioning traffic engineering in software defined wide area networks," in *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2023, pp. 596–601.

[164] K. Marszałek and A. Domański, "A fluid flow model for the software defined wide area networks analysis," *Scientific Reports*, vol. 15, p. 3713, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-88162-6

[165] Z. Guo, C. Li, Y. Li, S. Dou, B. Zhang, and W. Wu, "Maintaining the network performance of software-defined wans with efficient critical routing," *IEEE Transactions on Network and Service Management*, 2023.

[166] A. Navarro, R. Canonico, and A. Botta, "Software defined wide area networks: Current challenges and future perspectives," in *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*. IEEE, 2023, pp. 350–353.

[167] A. Botta, R. Canonico, A. Navarro, G. Stanco, and G. Ventre, "Adaptive overlay selection at the sd-wan edges: A reinforcement learning

approach with networked agents," *Computer Networks*, vol. 243, p. 110310, 2024.

[168] M. A. Ouamri, G. Barb, D. Singh, and F. Alexa, "Load balancing optimization in software-defined wide area networking (sd-wan) using deep reinforcement learning," in *2022 International Symposium on Electronics and Telecommunications (ISETC)*, 2022, pp. 1–6.

[169] J. Sommers, P. Barford, and W. Willinger, "A proposed framework for calibration of available bandwidth estimation tools," in *11th IEEE Symposium on Computers and Communications (ISCC'06)*. IEEE, 2006, pp. 709–718.

[170] A. Tootoonchian, M. Ghobadi, and Y. Ganjali, "Opentm: traffic matrix estimator for openflow networks," in *International Conference on Passive and Active Network Measurement*. Springer, 2010, pp. 201–210.

[171] T. Benson, A. Anand, A. Akella, and M. Zhang, "Microte: Fine grained traffic engineering for data centers," in *Proceedings of the seventh conference on emerging networking experiments and technologies*, 2011, pp. 1–12.

[172] M. Yu, L. Jose, and R. Miao, "Software {Defined}{Traffic} measurement with {OpenSketch}," in *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, 2013, pp. 29–42.

[173] S. R. Chowdhury, M. F. Bari, R. Ahmed, and R. Boutaba, "Payless: A low cost network monitoring framework for software defined networks," in *2014 IEEE Network Operations and Management Symposium (NOMS)*. IEEE, 2014, pp. 1–9.

[174] C. D. W. F. J. Carter, "Opensample: A low-latency, sampling-based measurement platform for sdn," *IBM Resear Division*, 2014.

[175] D. Katz and D. Ward, "Bidirectional forwarding detection (bfd)," Tech. Rep., 2010.

[176] G. Mirsky, G. Jun, H. Nydell, and R. Foote, "Rfc 8762: Simple two-way active measurement protocol," 2020.

[177] B. Yuan and C. Zhao, "Research on transmission delay of sd-wan cpe," in *2020 IEEE 20th International Conference on Communication Technology (ICCT)*. IEEE, 2020, pp. 955–960.

[178] R. Subramanian and J. Puthenparambil, "Shared memory enabled service plane optimization," in *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2020, pp. 683–684.

[179] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, 2003, pp. 39–44.

[180] A. Tirumala, "Iperf: The tcp/udp bandwidth measurement tool," *http://dast. nlanr. net/Projects/Iperf/*, 1999.

[181] S. Avallone, S. Guadagno, D. Emma, A. Pescape, and G. Ventre, "D-itg distributed internet traffic generator," in *First International Conference on the Quantitative Evaluation of Systems, 2004. QEST 2004. Proceedings.* IEEE, 2004, pp. 316–317.

[182] Ostinato, Accessed: 2025-01-09. [Online]. Available: http://ostinato.org/

[183] S. Troia, G. Maier, and S. Bregni, "Experimental evaluation of sd-wan performance in a municipal network test bed," in *2023 IEEE Latin-American Conference on Communications (LATINCOM)*, 2023, pp. 1–5.

[184] P. Iddalagi and A. Mishra, "Impact analysis of tunnel probing protocol on sd-wan's mainstream traffic," in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2023, pp. 252–259.

[185] A. Botta, R. Canonico, A. Navarro, G. Stanco, and G. Ventre, "Towards a highly-available sd-wan: Rapid failover based on bfd protocol," in *2023 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2023, pp. 153–158.

[186] P. Iddalagi and A. Mishra, "Reinforcement learning-based deep queue network model for tunnel health probing protocol management in sd-wan," in *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*. IEEE, 2024, pp. 467–473.

[187] 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*, 2023.

[188] RouteViews, Accessed: 2025-01-09. [Online]. Available: http://www.routeviews.org/

[189] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of the American statistical association*, vol. 91, no. 433, pp. 365–377, 1996.

[190] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, "Fast accurate computation of large-scale ip traffic matrices from link loads," *ACM SIGMETRICS Performance Evaluation Review*, vol. 31, no. 1, pp. 206–217, 2003.

[191] A. Soule, A. Nucci, R. L. Cruz, E. Leonardi, and N. Taft, "Estimating dynamic traffic matrices by using viable routing changes," *IEEE/ACM transactions on Networking*, vol. 15, no. 3, pp. 485–498, 2007.

[192] B. Melander, M. Bjorkman, and P. Gunningberg, "A new end-to-end probing and analysis method for estimating bandwidth bottlenecks," in *Globecom'00-IEEE. Global Telecommunications Conference. Conference Record (Cat. No. 00CH37137)*, vol. 1. IEEE, 2000, pp. 415–420.

[193] M. Jain, "Pathload: A measurement tool for end-to-end available bandwidth," in *Proc. of Passive and Active Measurements (PAM) Workshop, Mar. 2002*, 2002.

[194] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathchirp: Efficient available bandwidth estimation for network paths," in *Passive and active measurement workshop*, 2003.

[195] V. Mohan, Y. J. Reddy, and K. Kalpana, "Active and passive network measurements: a survey," *International Journal of Computer Science and Information Technologies*, vol. 2, no. 4, pp. 1372–1385, 2011.

[196] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and S. C. Diot, "Packet-level traffic measurements from the sprint ip backbone," *IEEE network*, vol. 17, no. 6, pp. 6–16, 2003.

[197] C. Yu, C. Lumezanu, Y. Zhang, V. Singh, G. Jiang, and H. V. Madhyastha, "Flowsense: Monitoring network utilization with zero measurement cost," in *Passive and Active Measurement: 14th International Conference, PAM 2013, Hong Kong, China, March 18-19, 2013. Proceedings 14*. Springer, 2013, pp. 31–41.

[198] eBPF, Accessed: 2025-01-09. [Online]. Available: https://ebpf.io/

[199] S. Sundberg, A. Brunstrom, S. Ferlin-Reiter, T. Høiland-Jørgensen, and J. D. Brouer, "Passive monitoring of network latency at high line rates," in *17th Swedish National Computer Networking Workshop (SNCNW 2022), Stockholm, June 16-17 2022*, 2022.

[200] S. Troia, M. Mazzara, L. M. M. Zorello, and G. Maier, "Performance evaluation of overlay networking for delay-sensitive services in sd-wan," in *2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*. IEEE, 2021, pp. 150–155.

[201] N. L. Van Adrichem, C. Doerr, and F. A. Kuipers, "Opennetmon: Network monitoring in openflow software-defined networks," in *Network Operations and Management Symposium (NOMS), 2014 IEEE*. IEEE, 2014, pp. 1–8.

[202] M. A. Makarem, A. Q. Algaolahi, A. A. Alhajri, and M. A. Razaz, "Enhancing load balancing & fault tolerance in real-time applications using sd-wan," in *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. IEEE, 2024, pp. 1–6.

[203] Redis, Accessed: 2025-01-09. [Online]. Available: https://redis.io/

[204] A. Navarro, R. Canonico, and A. Botta, "Software defined wide area networks: Current challenges and future perspectives," in *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*, 2023, pp. 350–353.

[205] R. Kundel, C. Gärtner, M. Luthra, S. Bhowmik, and B. Koldehofe, "Flexible content-based publish/subscribe over programmable data planes," in *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1–5.

[206] B. Sharma and D. Nadig, "ebpf-enhanced complete observability solution for cloud-native microservices," in *ICC 2024 - IEEE International Conference on Communications*, 2024, pp. 1980–1985.

[207] H. Yelland, "Compliance and regulatory considerations for sd-wan in manufacturing," Netify, 2024, Accessed: 2025-01-09. [Online]. Available: https://www.netify.com/learning/compliance-and-regulatory-considerations-for-sd-wan-in-manufacturing/

[208] "Versa networks achieves ipv6 certification, ensuring readiness for next-generation connectivity, interoperability, and security," Versa Networks, 2023, Accessed: 2025-01-09. [Online]. Available: https://versa-networks.com/news/2023/versa-networks-achieves-ipv6-certification-ensuring-readiness-for-next-generation-connectivity-interoperability-and-security/

[209] "Unmasking single points of failure: Vulnerabilities in centralized databases," Inery Blog, 2024, Accessed: 2025-01-09. [Online]. Available: https://inery.io/blog/article/unmasking-single-points-of-failure-vulnerabilities-in-centralized-databases/

[210] R. G. Clegg, M. S. Withall, A. W. Moore, I. W. Phillips, D. J. Parish, M. Rio, R. Landa, H. Haddadi, K. Kyriakopoulos, J. Auge *et al.*, "Challenges in the capture and dissemination of measurements from high-speed networks," *IET communications*, vol. 3, no. 6, pp. 957–966, 2009.

[211] "How ai and machine learning are revolutionizing network management," Netop Blog, Accessed: 2025-01-09. [Online]. Available: https://netop.cloud/blog/how-ai-and-machine-learning-are-revolutionizing-network-management/

[212] M. Yu, "Network telemetry: towards a top-down approach," *ACM SIGCOMM Comput. Commun, Rev. 49*, 2019. [Online]. Available: http://dx.doi.org/10.1109/SECPRI.2003.1199330

[213] C. Kim, A. Sivaraman, N. Katta, A. Bas, A. Dixit, and L. Wobker, "In-band network telemetry via programmable dataplanes," 2015. [Online]. Available: https://nkatta.github.io/papers/int-demo.pdf

[214] T. P and A. W. and Group and et al., ""in-band network telemetry (int) dataplane specification - version 2.1"," 2020. [Online]. Available: https://p4.org/p4-spec/docs/INT_v2_1.pdf

[215] L. Tan, W. Su, W. Zhang, J. Lv, Z. Zhang, J. Miao, X. Liu, and N. Li, "In-band network telemetry: A survey," *Computer Networks*, vol. 186, p. 107763, 2021.

[216] C. Min, D. Zhao, and H. Lu, "The processing method of the message based on the in-band network telemetry technology," in *2022 International Conference on Service Science (ICSS)*. IEEE, 2022, pp. 21–24.

[217] S. Troia, E. Gregorini, J. P. Asdikian, M. Li, and G. Maier, "Real-time delay measurement in sd-wan based on in-band network telemetry," in *2024 IEEE 25th International Conference on High Performance Switching and Routing (HPSR)*, 2024, pp. 149–154.

[218] "Ping [online]," "https://linux.die.net/man/8/ping", 2023.

[219] "Traceroute [online]," "https://linux.die.net/man/8/traceroute", 2023.

[220] E. H. Halili, *Apache JMeter*, 2008.

[221] Dig, Accessed: 2025-01-09. [Online]. Available: https://linux.die.net/man/1/dig

[222] M. Mortimer, "iperf3 documentation," 2018, Accessed: 2025-02-06. [Online]. Available: https://readthedocs.org/projects/iperf3-python/downloads/pdf/latest/

[223] TRex, Accessed: 2025-01-09. [Online]. Available: https://trex-tgn.cisco.com/

[224] N. Pinaeva and V. Antonenko, "Warp: Wan connection bandwidth monitoring and prediction," in *2020 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. IEEE, 2020, pp. 228–234.

[225] V. Tran, J. Tourrilhes, K. Ramakrishnan, and P. Sharma, "Accurate available bandwidth measurement with packet batching mitigation for high speed networks," in *2021 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. IEEE, 2021, pp. 1–6.

[226] R. Prasad, M. Jain, and C. Dovrolis, "Effects of interrupt coalescence on network measurements," in *International Workshop on Passive and Active Network Measurement*. Springer, 2004, pp. 247–256.

[227] Q. Yin, J. Kaur, and F. D. Smith, "Can bandwidth estimation tackle noise at ultra-high speeds?" in *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 2014, pp. 107–118.

[228] Z. Guo, L. Qi, S. Dou, J. Weng, X. Fu, and Y. Xia, "Maintaining control resiliency for traffic engineering in sd-wans," *IEEE/ACM Transactions on Networking*, 2024.

[229] S. Dou and Z. Guo, "Maintaining predictable traffic engineering performance under controller failures for software-defined wans," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2025.

[230] Z. Guo, S. Dou, and W. Jiang, "Improving the path programmability for software-defined wans under multiple controller failures," in *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*. IEEE, 2020, pp. 1–10.

[231] S. Dou, Z. Guo, and Y. Xia, "Programmabilitymedic: Predictable path programmability recovery under multiple controller failures in sd-wans," in *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2021, pp. 461–471.

[232] S. Dou and Z. Guo, "Path programmability recovery under controller failures for sd-wans: Recent advances and future research challenges," *IEEE Communications Magazine*, 2024.

[233] Z. Guo, S. Dou, W. Wu, and Y. Xia, "Toward flexible and predictable path programmability recovery under multiple controller failures in software-defined wans," *IEEE/ACM Transactions on Networking*, 2023.

[234] K. Basu, A. Hamdullah, and F. Ball, "Architecture of a cloud-based fault-tolerant control platform for improving the qos of social multimedia applications on sd-wan," in *2020 13th International Conference on Communications (COMM)*. IEEE, 2020, pp. 495–500.

[235] Z. Guo, S. Dou, W. Jiang, and Y. Xia, "Toward improved path programmability recovery for software-defined wans under multiple controller failures," *IEEE/ACM Transactions on Networking*, 2023.

[236] L. Borgianni, C. Bua, S. Ghadir, D. Ghadir, D. Adami, and S. Giordano, "Enhancing wildlife monitoring in variable network coverage areas with deep rl-based sd-wans," in *2024 IEEE 29th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2024, pp. 1–6.

[237] S. Sun, M. Kiran, and W. Ren, "Mamrl: Exploiting multi-agent meta reinforcement learning in wan traffic engineering," *arXiv preprint arXiv:2111.15087*, 2021.

[238] S. Troia, L. M. M. Zorello, and G. Maier, "Sd-wan: how the control of the network can be shifted from core to edge," in *2021 International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, 2021, pp. 1–3.

[239] X. Chen, Y. Zhang, C. Jiang, C. Xu, Z. Yuan, and G.-M. Muntean, "Revenue-oriented optimal service offloading based on fog-cloud collaboration in sd-wan enabled manufacturing networks," *IEEE Transactions on Network Science and Engineering*, 2025.

[240] Y. Zhang, C. Xu, and G.-M. Muntean, "A novel distributed data backup and recovery method for software defined-wan controllers," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 01–06.

[241] I. O. Adebayo, M. O. Adigun, and P. Mudali, "Neighbourhood centality based algorithms for switch-to-controller allocation in sd-wans," in *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, 2023, pp. 1–6.

[242] F. Altheide, S. Buttgereit, M. Rossberg, and G. Schaefer, "Increasing resilience of sd-wan by distributing the control plane," in *2023 14th International Conference on Network of the Future (NoF)*. IEEE, 2023, pp. 10–18.

[243] F. Altheide, S. Buttgereit, and M. Rossberg, "Increasing resilience of sd-wan by distributing the control plane [extended version]," *IEEE Transactions on Network and Service Management*, 2024.

[244] B. Yuan, H. Li, and H. Yu, "A study of selection mechanisms for hub cpes in sd-wan network systems," in *2024 5th Information Communication Technologies Conference (ICTC)*. IEEE, 2024, pp. 117–121.

[245] C.-Y. Chang, T. G. Ruiz, F. Paolucci, M. A. Jiménez, J. Sacido, C. Papagianni, F. Ubaldi, D. Scano, M. Gharbaoui, A. Giorgetti *et al.*, "Performance isolation for network slices in industry 4.0: The 5growth approach," *IEEE Access*, vol. 9, pp. 166 990–167 003, 2021.

[246] A. Giorgetti, D. Scano, and L. Valcarenghi, "Guaranteeing slice performance isolation with sdn," *IEEE Communications Letters*, vol. 25, no. 11, pp. 3537–3541, 2021.

[247] L. Borgianni, D. Adami, and S. Giordano, "Optimizing network performance and reliability with an integrated sd-wan and satellite 6g architecture," in *2023 2nd International Conference on 6G Networking (6GNet)*, 2023, pp. 1–4.

[248] L. Qi, S. Dou, Z. Guo, C. Li, Y. Li, and T. Zhu, "Low control latency sd-wans for metaverse," in *2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 2022, pp. 266–271.

[249] K. Duncan, D. Hoenigmann, and T. Guyah, "Network communications evolution's effects on air traffic systems–nuisance or national risk," in *2023 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. IEEE, 2023, pp. 1–8.

[250] L. Borgianni, S. Giordano, and C. Chafe, "Democratizing networked music performance with a rl-based sd-wan," *IEEE Communications Magazine*, vol. 62, no. 12, pp. 26–32, 2024.

[251] "Report: Generative ai taking over sd-wan management," Campus Technology, 2024, Accessed: 2025-01-09. [Online]. Available: https://campustechnology.com/Articles/2024/10/04/Report-Generative-AI-Taking-Over-SD-WAN-Management.aspx

[252] "Mars relay network and the interplanetary internet," NASA Science, 2025, accessed: 2025-01-14. [Online]. Available: https://science.nasa.gov/planetary-science/programs/mars-exploration/mars-relay-network-interplanetary-internet/?utm_source

[253] "Generation interplanetary internet," SpaceRef, 2025, accessed: 2025-01-14. [Online]. Available: https://spaceref.com/uncategorized/generation-interplanetary-internet/

[254] H. Xiang, C. Yi, K. Wu, J. Chen, J. Cai, D. Niyato, and X. Shen, "Realizing immersive communications in human digital twin by edge computing empowered tactile internet: Visions and case study," *IEEE Network*, pp. 1–1, 2024.

[255] J. S. Zhihan Lyu, James J. Park and H. Song, "State-of-the-art human-computer-interaction in metaverse," *International Journal of Human–Computer Interaction*, vol. 40, no. 21, pp. 6685–6689, 2024. [Online]. Available: https://doi.org/10.1080/10447318.2024.2405565