Data center interconnection for the cloud era

Enabling a scalable, dynamic cloud ecosystem with network automation

White paper

The global data center market is experiencing huge demand driven by enterprise digital transformation, changes to working practices and applications, services and workloads migrating to the cloud. The cloud ecosystem and the data centers that form it require highly interconnected network infrastructure to deliver exceptional experiences for customers regardless of their location. This highlights the need for global and regional networks that connect and interconnect data centers in a scalable, secure and reliable way.

This white paper explains the need for multi-layer data center interconnection networks and how they need to support dynamic access to cloud applications and services. Today's multiple, costly, static networks require manual provisioning and intervention across multiple layers and domains. They need to evolve to enable an automated, multi-layer data center interconnection fabric that supports dynamic connectivity to cloud application and services running in both edge and core data centers.

Contents

Introduction

The global data center market is experiencing huge demand driven by the cloud, enterprise digital transformation and changes to working practices. Enterprises are accelerating their digital transformations by continuing to migrate applications, services and workloads to the cloud. This increases the demand for data centers that form the cloud.

Both consumers and enterprises continue to generate huge volumes of data, which is a key metric in the growth of the cloud and data center infrastructure spending. According to analyst firm Dell'Oro, CAPEX spending on data center infrastructure will reach $377 billion US dollars by 20261 , with the hyperscale cloud providers accounting for more than half the market total.

Although most of this data is stored in data centers owned by cloud providers, a significant proportion of enterprise data is stored off-premises in data centers owned by colocation providers, hosting providers and network operators. These companies continue to invest in their data center facilities to cope with the growth of enterprise data and the demand for access to cloud applications and services.

The cloud and the data centers that make up the cloud form an ecosystem that enables digital transformation and the digital economy. This cloud ecosystem requires a highly interconnected data center infrastructure to deliver exceptional experiences for customers regardless of their location

or where the applications and services are running. As a result, there's a need for global and regional networks that connect and interconnect data centers in a scalable, secure and reliable way.

Equinix's most recent Global Interconnection Index2 and TeleGeography's Internet Map3 forecast a compound annual growth rate of 44 percent for global data center interconnection and predict that interconnection bandwidth will be 15 times larger than Internet bandwidth by 2024.

To support this rapidly changing environment, the data center interconnection networks of service providers, cloud providers, and carrier-neutral and colocation data center providers need a profound shift to support new cloud applications and services regardless of where they are located. 5G, the Internet of Things (IoT), artificial intelligence (AI), augmented reality (AR) and Industry 4.0 all bring explosive growth in traffic volume, with increasingly dynamic and elastic traffic demands. As a result, data center interconnection networks require greater scalability, resiliency and automation.

The cloud changes how we need to connect data centers

The bandwidth, latency and reliability requirements of cloud applications and services determine where they should be located and processed. Real-time and latency-sensitive applications and services are best located in edge clouds that are closer to end users, while non-real-time applications and services can remain in the core cloud.

These varying requirements and evolving edge cloud infrastructures impact current network architecture, and the traditional distinction between metro, aggregation and core networks is disappearing. Today's

1 Hyperscalers to Lift Data Center Capex to $377 Billion in 2026, Dell'Oro, August 2022
2 The Global Interconnection Index (GXI) Volume 5, Equinix Q2, 2021
3 Global Internet Map 2021, TeleGeography, 2021

costly, multiple, static networks—which require manual provisioning and intervention across multiple layers and domains—must evolve rapidly to become a single, automated, multi-layer network interconnection fabric that supports dynamic connectivity between edge and core data centers.

Rapid automation and efficient optimization are critical as new applications emerge, server virtualization increases, data centers become more distributed and edge cloud becomes more prevalent. The ability to interconnect data centers using any form of intermediate infrastructure—such as lit or dark fiber, managed wavelengths or multi-technology paths (e.g., IP/MPLS, segment routing, VPN, OTN or Ethernet)—is also becoming critical.

Organizations are increasingly seeing the benefit of using third-party edge cloud infrastructure to help them implement services and applications closer to their customers, partners and end users. Examples include:

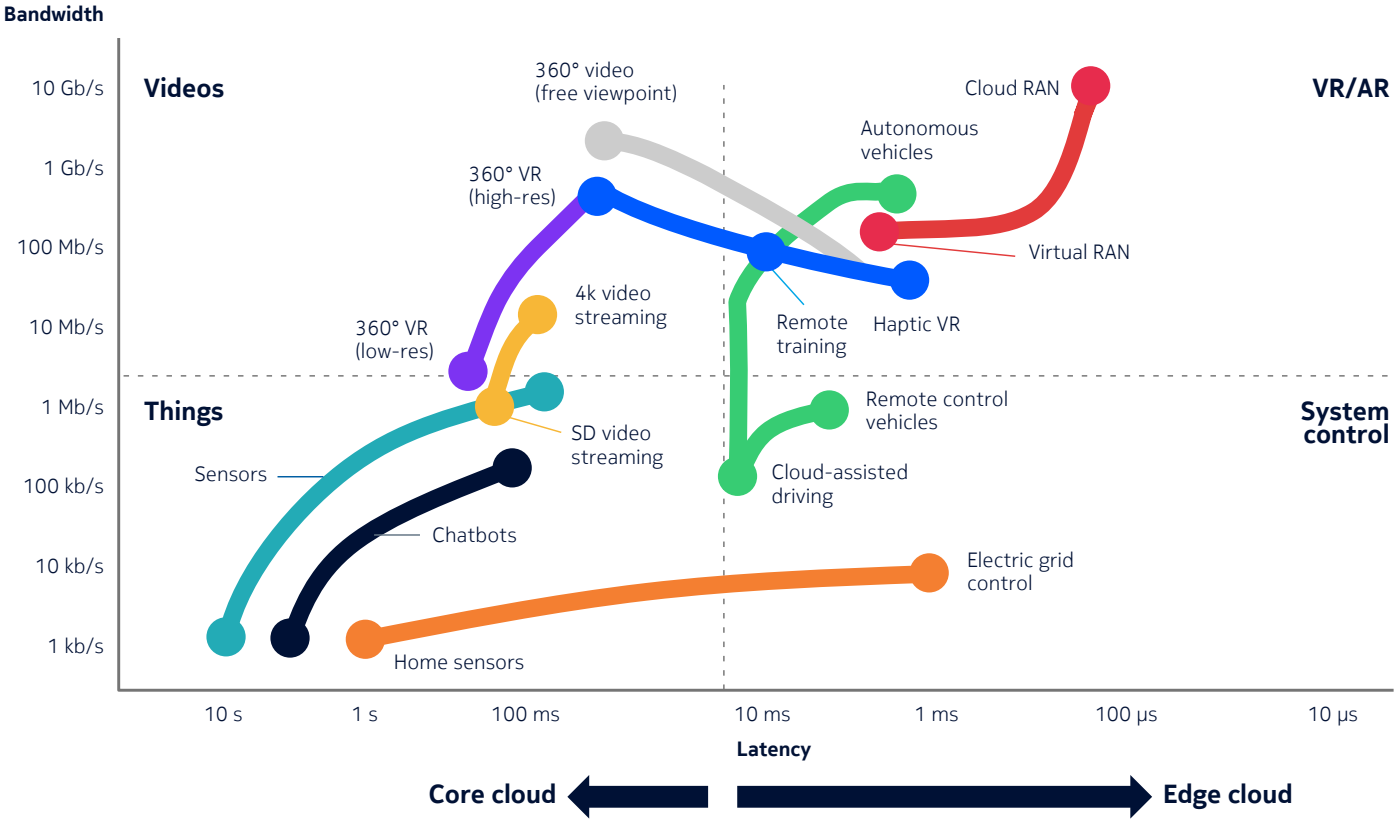
- Enterprises undergoing digital transformation to support new applications, such as IoT, AI, machine-to-machine and Industry 4.0, as well as disaster recovery and business continuity
- Cloud providers seeking to use interconnection to expand the footprint and reach of their applications and services into new markets
- Network operators and carrier providers that offer wholesale network connectivity services to connect customers to applications and services in cloud data centers
- Data center and interconnection providers implementing platforms and portals that provide their customers with dynamic access to the cloud ecosystem
- Mobile operators looking to extend 5G network rollouts more cost effectively by deploying cloud-based infrastructure.

Carrier-neutral and colocation data center providers that have multi-tenant facilities and operate in key markets globally are ideally placed to offer computing, networking and interconnection to support organizations as they implement edge and core cloud applications and services. Multi-tenant cloud computing is also driving the need for dynamically instantiated virtual networks to enable these providers to interconnect their data centers and better serve the multi-cloud needs of various tenant organizations.

Edge cloud introduces new architecture challenges

The introduction of 5G, IoT, AI, VR/AR and Industry 4.0 promises to support a range of new applications and services. However, many of these new applications and services have specific bandwidth and ultra-low latency requirements, as shown in Figure 1.

Figure 1. Caption goes above the figures and tables



A new network architecture is needed to deploy very large-scale edge cloud infrastructure with the required bandwidth and latency characteristics to support these new applications. This network architecture is driven by the needs of applications and where they should be located and processed to meet these bandwidth, latency and reliability requirements. For example:

- Enhanced 5G mobile broadband applications requiring high bandwidth but best-effort service can be processed in core or near-edge data centers depending on requirements.
- Applications that require ultra-reliable, ultra-low latency with redundant paths, such as robotic and autonomous vehicles, are best processed in edge data centers.
- Dynamic applications that require massive connectivity and signaling but low bandwidth, such as IoT applications, are best processed in core data centers.
- Some applications, such as AI, will need additional compute and network resources, and elements of AI may need to be processed in both edge and core data centers.

These applications are creating additional compute and networking demands at both the edge and core, requiring a new approach that ties edge to core in a much more scalable, agile and dynamic way. Higher levels of network resiliency and performance are also needed as more applications and services are deployed and more devices are connected. In addition, the network must be able to cope with the increasing number, size, sophistication and frequency of malicious cyber attacks.

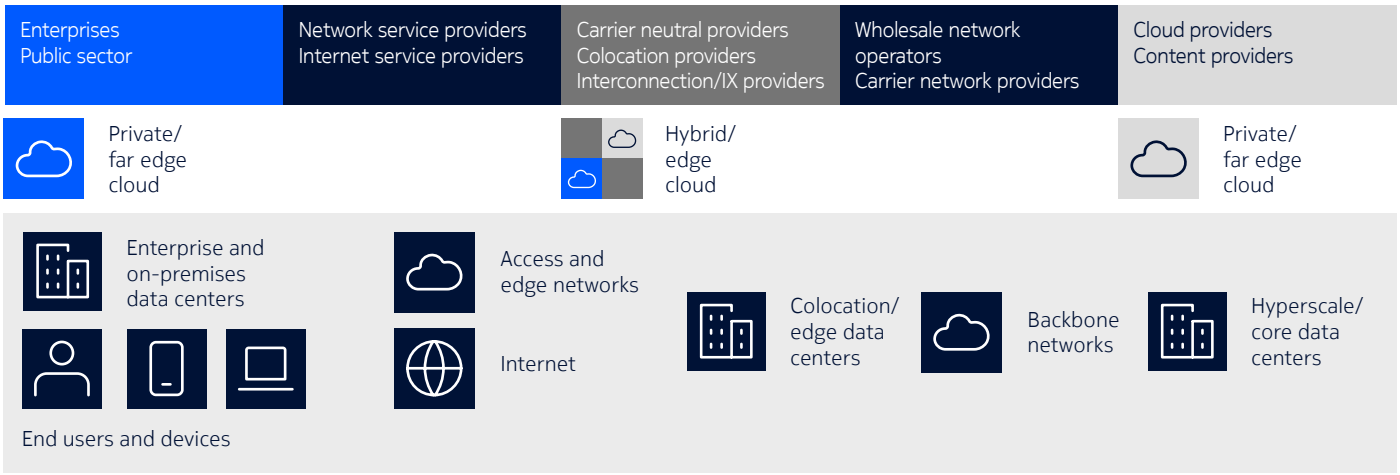
The challenge is that although today's infrastructure is reliable, resilient and secure, it is enabled by complex, multi-layer networks, with little or no integration between layers. Typically, this means that each layer is over-engineered to provide the reliability, resilience and security required to accommodate highly variable traffic demands.

Network infrastructure needs to be more scalable and dynamic and must also reduce operating costs and increase efficiency. It must also capture new revenue-generating opportunities presented by new edge cloud applications and continue to support traditional business models.

The challenges of data center connectivity in the cloud era

The many participants in the cloud ecosystem (see Figure 2) include cloud providers, communications service providers, carrier-neutral network operators, data center operators, colocation and hosting providers, and interconnection providers. There are also the enterprises, public sector organizations and consumers who use cloud applications and services.

Figure 2. Participants in the cloud ecosystem



Providers not only need to consider the impact of the cloud on their data centers, they must also consider the impact on the networks that connect and interconnect them. While there are many opportunities to offer retail and wholesale data center connectivity, operators must support deterministic end-to-end services such as wholesale transport, data center interconnect (DCI) and IP connectivity for multiple customers, who all have their own requirements for speed, latency, availability and security.

There are three key networking capabilities that operators need to consider for data center connectivity in the cloud: connection, interconnection and automation.

Connection

- Connect data centers to maximize fiber capacity and reach while containing costs.
- Understand key fiber route trends, both terrestrial and subsea.
- Plan for greater network resiliency through regional and international route diversity.

Operators can meet connection challenges by deploying agile, scalable and resilient optical mesh networks and point-to-point DCI solutions that maximize fiber capacity and reach at the lowest cost per bit.

Interconnection

- Interconnect data centers with multiple service providers, cloud providers and internet providers.
- Create a scalable, secure and reliable cloud interconnection platform.
- Give customers more flexibility and control to access cloud services on demand.

Operators can meet these challenges by implementing secure peering, data center gateway and IP routing solutions that offer comprehensive IP edge routing features while providing IP core routing that is optimized for high performance.

Automation

- Control costs and increase efficiency by automating network operations.
- Simplify and integrate multi-layer networks to reduce complexity.
- Program the network to meet the dynamic requirements of applications and workloads.

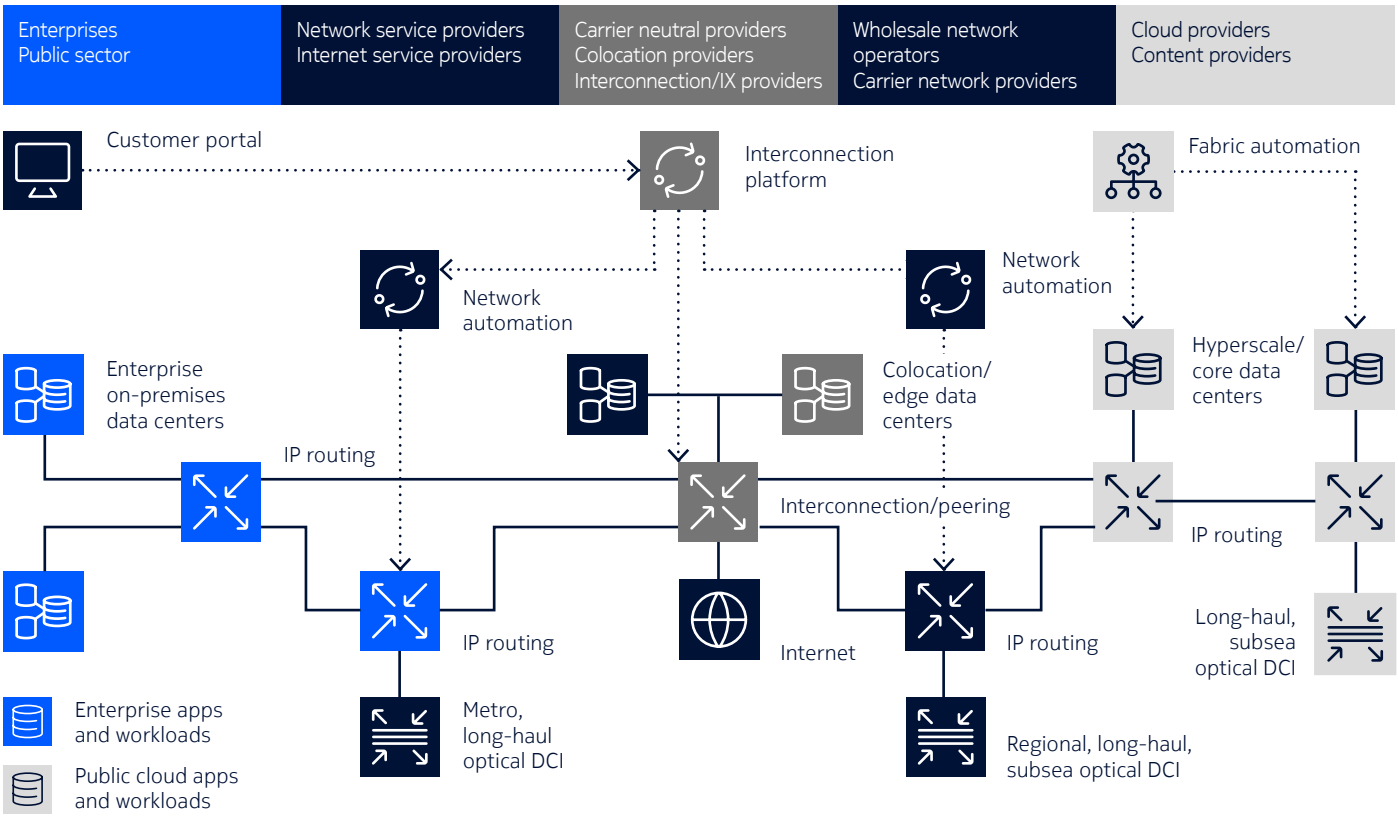
Operators can meet these challenges by deploying network automation to simplify operations, responding quickly to fast-changing demand, and optimizing network resources to ensure maximum service performance and reliability.

Automation of data center connection and interconnection is critical to enabling a dynamic cloud ecosystem and provides the following benefits for providers and customers:

- Enables faster delivery of cloud applications and services. Network connectivity, bandwidth and services can be deployed in minutes to meet customer demand rather than the days or weeks that traditional network configuration, provisioning and operations processes can take.
- Reduces network operations complexity and cost. Automating the configuration, provisioning, deployment and operation of data center connectivity through a “single pane of glass” operations approach not only reduces network operations costs but helps maximize the use of network resources and assets.
- Offers customers access to more cloud services and applications. By developing interconnection platforms, operators offer customers access to an ecosystem of cloud applications and services wherever they are located and enable them to distribute workloads where and when they need them.
- Gives customers more control. Using interactive web portals, customers can also have more control of how they connect to cloud services and deploy new applications. Web portals allow customers to selfprovision cloud services and applications without the need to configure complex network connectivity; this makes the network invisible to this process and cloud connectivity as easy to consume as cloud compute and storage.

Figure 2. Participants in the cloud ecosystem

Figure 3. Automated data center interconnection



Connecting data centers in the cloud

Connecting data centers with optical networks provides DCI and transport services at scale, capacity and at the lowest cost.

The cloud requires very large amounts of data to be transported between data centers. It also requires replication of this data using synchronous and asynchronous backup between primary, secondary and tertiary data centers, both for on- and off-premises business continuity and disaster recovery. Often, all of these services also need to be provided between regions.

In addition, cloud providers, retail service providers and regional network operators need wholesale transport services to extend footprint and reach, gain access to regional data center hubs, and provide connectivity between edge and core data centers. Each has specific requirements for speed, capacity, latency, availability and security.

Wholesale transport and data center interconnect services must support deterministic end-to-end services with guaranteed Service Level Agreements (SLAs) over a secure, shared network infrastructure that may comprise point-to-point, ring and mesh topologies. Solutions that provide the capacity to meet current needs but can also scale incrementally to meet future needs without major upgrades help operators who are moving from 10G, 40G or 100G transport links to 100G, 400G and higher transport links.

Operators are under pressure to maximize the utilization and efficiency of valuable fiber assets, contain operations costs and improve operations efficiency. Technologies such as coherent optics, digital signal processing and the associated algorithms are key to scaling optical network capacity and improving efficiency.

Interconnecting data centers in the cloud ecosystem

Interconnection routes traffic between multiple networks, connecting customers with cloud services, applications and workloads wherever they are located. It involves routing traffic from one network to another and between different operators' data centers in a process known as peering.

Peering enables two networks to connect and exchange traffic directly without needing to pay a third party to carry traffic across the internet. In some cases, it involves multiple networks, each peering with the other to pass traffic from the source to destination; for example, from a customer office to an application running in a core cloud data center.

Public peering provided by an internet exchange provider enables one network to exchange traffic with multiple other networks. Once connected to a public peering exchange, operators can set up or remove connections to other operators' networks without needing to physically re-provision circuits. Private peering, typically between only two parties, enables one network to exchange traffic with another network; for example, either physically using a fiber connection or virtually using a VLAN.

Peering typically takes place in carrier-neutral data centers, where many participants in the cloud ecosystem are collocated. These richly connected data centers are valuable assets and are the ideal location for interconnection, whereby any party can connect with any other collocated party to gain access to cloud services and applications—wherever these are located.

This interconnection ecosystem is enhanced by automating network connectivity and peering, and by using web portals to make it easy for customers to provision connectivity to cloud applications and services or to connect to partners, suppliers and other enterprises in private ecosystems.

Network automation for the cloud ecosystem

Today's networks are static, multi-layered and costly, and need to be more integrated and automated to offer flexible and resilient connectivity between edge and core data centers and more dynamic access to the cloud. Automation of data center connection and interconnection is critical for this and needs to be optimized and efficient, with the ability to use multiple infrastructures, such as illuminated or dark fiber, managed wavelengths and multi-technology IP/optical networks.

The challenge is that while today's network infrastructure is reliable, resilient and secure, it is enabled by IP and optical layers with little or no integration between them. Typically, this means that each layer is over-engineered and under-utilized for the actual traffic demands.

IP/optical network automation helps operators to increase operations efficiency and reduce operations costs. It also makes the network more responsive to customer demand. Customers want dynamic access to cloud-based applications and services that are located in multiple data centers and the ability to distribute cloud-based workloads between data centers where and when needed.

To achieve these goals, the underlying network must be transparent and able to respond to demand dynamically. Increasingly, customers want cloud-to-cloud connectivity and access to multi-clouds that

enable them to distribute or move workloads dynamically, and access distributed applications and cloud services when and where they need them.

Giving customers more control

Network automation puts more control in the hands of customers and responds to their needs within minutes or hours rather than days or weeks. It is key to enabling more dynamic access to cloud applications and services by providing software-programmable IP/optical data center connectivity and interconnection that is customizable through APIs.

By integrating customer web portals with network interconnection platforms, operators can automate the configuration, provisioning and deployment of network resources to make the whole process as seamless as possible.

Customers have more control to bring up applications, access services and spin up workloads when and where needed to serve their needs in any location. They don't need to contact the network operator and ask for more capacity in a specific data center or between data centers. Benefits include:

- Flexible, pay-per-use connectivity and on-demand bandwidth
- Agile provisioning of cloud applications and services
- Direct access to network, service, cloud and interconnection providers
- Secure connectivity for multi-cloud and cloud-to-cloud.

Helping operators expand their role

Network automation enables operators to reduce costs, increase efficiency and focus on deploying applications and services for customers rather than configuring complex network connectivity. It speeds up service delivery and helps to improve customer retention and improve the customer experience.

Network automation also helps operators to collaborate, develop partnerships and joint ventures, and participate more fully in the cloud ecosystem. Network operators can partner with cloud providers to offer optical transport and data center interconnect services. Collocation and data center operators can partner with network operators to offer wholesale network connectivity services. Wholesale network operators can offer fiber, wavelength, optical transport and IP transit services to retail service providers. Regional service providers can partner with global network operators to extend their network footprint and reach.

Technical features and benefits

Data center interconnection based on agile, flexible and high-performance IP/optical infrastructure enables a WAN fabric that can support edge cloud and data center connectivity, satisfy emerging requirements, and anticipate future application demands. Here are the main features and benefits of such a data center interconnection fabric.

Scalability and adaptability

The fabric provides highly scalable IP routing and optical transport with fully flexible traffic engineering based on dynamic traffic patterns. It provides a distributed interconnection architecture that can scale out horizontally without impacting services and reduces the number of ports required in an interconnected fabric.

Segment routing

The fabric can scale significantly because segment routing does not require any control plane signaling or any changes to the MPLS data plane. Segment routing allows the network architecture to be simplified because only edge devices need to maintain state information, removing this burden from core devices.

Granular traffic engineering

Path computation element (PCE) servers maintain network-wide segment routing topology, the state of active paths and reserved resources in a centralized traffic engineering database. Edge devices communicate with a PCE server to request path computation, and the software defined network (SDN) controller determines when and where to establish paths based on traffic and application performance requirements.

Performance and latency

The fabric supports very high traffic loads, is highly resilient and enables smooth network evolution; for example, from ring to mesh architectures. It supports applications with very low latency requirements and provides real-time control of ultra-low latency and dynamically adapting applications.

Reduced operations cost

The multi-layer IP/optical infrastructure helps to reduce operations costs. The optical layer provides highly scalable bandwidth and maximizes network capacity to provide efficient, ultra-low-cost transport. The IP layer provides highly scalable, reliable and high-performance routing using platforms based on feature and performance requirements at the edge or core of the network. Multi-layer SDN control and automation reduces complexity and helps to simplify operations and reduce costs.

Multilayer security

Network analytics with real-time monitoring and detection enables 360-degree visibility and protection against external and internal distributed denial of service (DDoS) attacks. Routers with programmable custom silicon provide line rate packet filtering for highly scalable mitigation and protection. Data encryption—such as AES 256-bit symmetric encryption at the optical layer and IPsec and Network Group Encryption at the IP layer—ensures a multi-level approach to network and data security.

Resiliency

The fabric provides ultra-high reliability and the ability to recover from multiple link failures using a combination of IP protection protocols and optical restoration protocols. Avoiding sets of routing links or paths sharing a common optical transport resource ensures high resiliency by avoiding all links in a set being affected if the common resource fails.

Multi-tenancy

The fabric enables partitioning of the physical infrastructure into multiple virtual infrastructures or network slices that are independently and automatically controlled with full autonomy. In this way, the network fabric can support multiple tenants or organizations simultaneously and dynamically.

Simplification and standardization

The fabric supports service abstraction with defined optimization constraints and objectives relevant to the application traffic. For easier onboarding of multiple tenants, the solution supports easy integration with existing network infrastructures. The fabric provides scalability, configuration consistency and the flexibility to leverage network programmability, enabling flexible service models.

Table 1 summarizes the key differences between a traditional network approach to connecting data centers to an approach based on data center interconnection fabric.

Table 1. Key differences between traditional network designs and next-generation network fabrics

Traditional approach	Fabric approach
Interconnected ring or point-to-point architectures	Dynamic mesh architecture
Separate IP and optical layers operating in separate network management silos	SDN programmable IP/optical network fabric with multi-layer automation
1:1 redundancy with self-healing protection	1:n redundancy with dynamic restoration
Manual and static traffic steering and network peering with low (<= 50%) resource utilization	Automated and dynamic traffic steering and network peering with high (>=50%) resource utilization
Traffic engineering based on RSVP-TE signaling with limited scalability and coarse control	Traffic engineering based on segment routing with high scalability and granular control
Security-unaware network requiring additional hardware investments to protect itself	360-degree security-aware network with built-in capabilities to protect itself and its users
Limited resiliency with single failure recovery (1:1 or 1+1 redundancy)	Dynamic resiliency that accommodates multi-link and multicomponent failure (N:M and N+M redundancy)
Complex, manual operations with limited visibility and control across network layers	Simple, automated operation using multi-layer SDN to coordinate IP and optical domains

Dynamic multi-layer resource optimization and protection based on SDN adaptive topology and traffic steering enables the fabric to mitigate the impact of traffic uncertainty. The integration of IP/optical control and management planes enables a more dynamic, resilient, cost-effective way to deliver highly scalable bandwidth and services on demand.

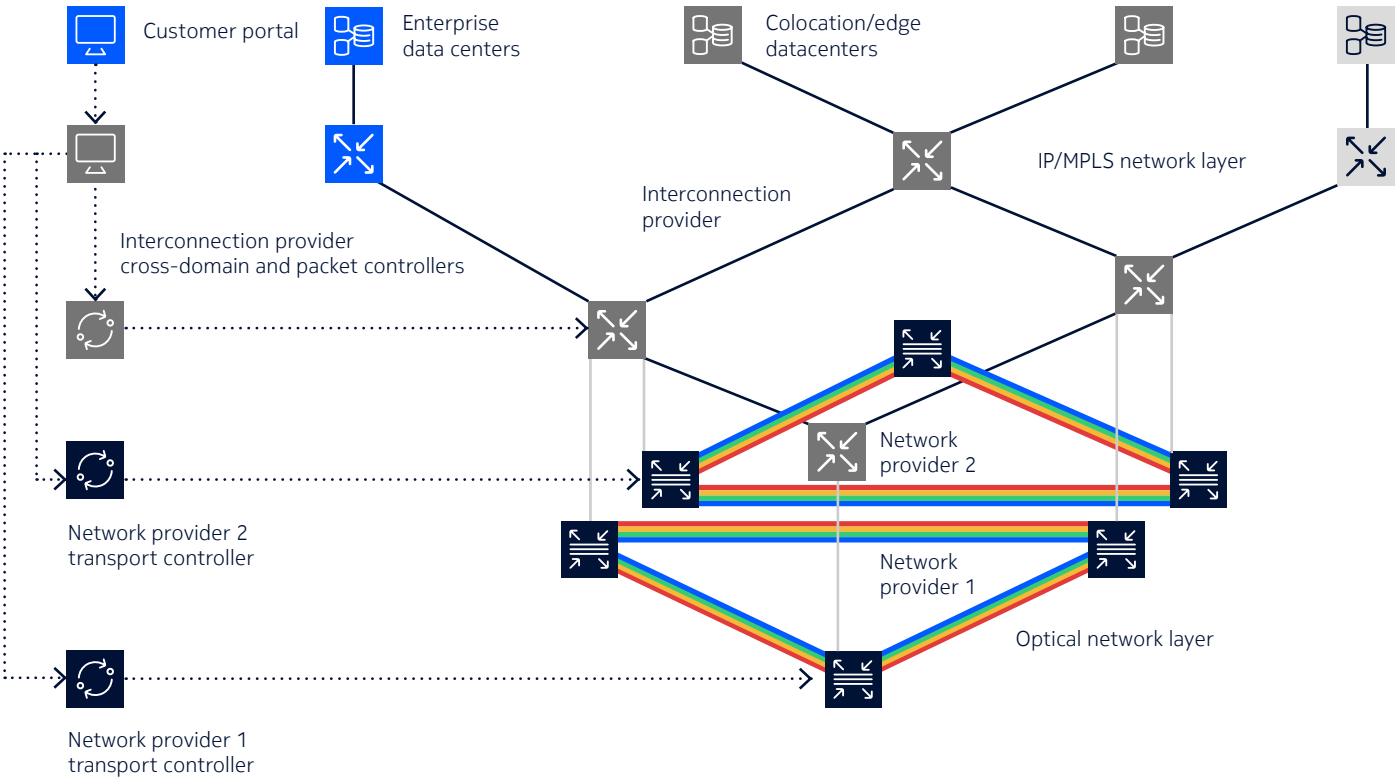
Multi-layer, cross-domain SDN control is provided by a cross-domain network resource controller that coordinates between transport controllers and IP/MPLS packet controllers, manages cross-domain Layer 0 – Layer 3 services, and optimizes resources, as shown in Figure 4.

The cross-domain controller provides intelligent resource control for multi-domain IP/optical networks. It dynamically creates optimal paths across multiple domains that are separated by IP/optical or vendor boundaries and instantiates hybrid IP/optical services in seconds by selecting the best path using available network resources.

The cross-domain controller is especially critical for hybrid IP/optical networks where multi-layer path stitching and provisioning is often a long and complex process.

Multi-layer optimization also ensures more efficient use of network resources than techniques that optimize the IP and optical layers independently.

Figure 4. Interconnection across multi-layer, multi-domain networks



Some examples of how the cross-domain controller provides awareness and coordination across IP and optical transport layers include:

- Optically disjoint IP routing: Allows the packet controller to use Shared Risk Link Group (SRLG) in path disjoint routing algorithms, ensuring that IP paths (such as an active and standby path) do not share a common optical path or fiber.
- Optical latency-aware IP routing: Enables a packet controller to use optical latency information to optimize the IP path based on latency constraints.

- Multi-layer simulation: Allows analysis of what-if scenarios to support network planning and troubleshooting with multi-layer visibility; for example, a scenario to identify all IP and IP/MPLS links that use a specific fiber to determine the impact of a maintenance event or failure.
- IP/optical maintenance coordination: Provides a proactive approach to moving IP traffic from optical links that are targeted for maintenance; this improves operational efficiency, reduces operational complexity and ensures no risk of service outage.
- IP/optical diversity analysis: Generates audit reports about diversity of link layer interconnects and passive analysis and correlation of optical topology and IP router ports.
- Common assurance: Provides a common list of IP/MPLS and optical alarms as well as correlation and root-cause analysis of alarms, with a common API for operations support system (OSS) integration.

The cross-domain controller integrates with platforms that orchestrate end-to-end services across hybrid (physical and virtual) resource domains. These platforms provide end-to-end service life cycle orchestration of complex hybrid services, including virtualized cloud resources and traditional physical resources in multi-vendor, multi-technology and multi-domain environments.

Typically, the platforms provide resource discovery, service design, service order, service orchestration and deployment, and work in conjunction with fulfilment systems in traditional OSSs.

Operational benefits

Data center interconnection using a single integrated IP/optical network and multi-layer SDN automation and control can help providers reduce the operational cost and improve the operational efficiency of both existing services and emerging cloud applications. Introducing an SDN-programmable IP/optical network offers the following operational benefits for interconnecting data centers:

- Multi-layer discovery and visibility that simplifies operations and reduces operational costs; for example, by:
  - Enabling easier troubleshooting across IP and optical layers
  - Coordinating maintenance without error-prone manual processes
  - Running what-if analysis to ensure IP traffic is never impacted if failures occur
  - Providing a Layer 0 (L0) GMPLS control plane for automated optical wavelength rerouting and restoration.
- Efficient use of IP and optical resources that increases network utilization and maximizes return on investment; for example, by:
  - Automating the setup of dynamic optical services with resiliency
  - Increasing traffic flow protection and utilization of router ports by pooling (ECMP)
  - Avoiding 1+1 resiliency by leveraging L0 GMPLS optical restoration for 1:N resiliency.
- Multi-layer traffic engineering to enable better service quality differentiation and improved resiliency and latency for new edge cloud applications; for example, by:
  - Ensuring optical link diversity for IP routing by using SRLG
  - Providing comprehensive correlation of network topology; for example, for latency
  - Implementing a multi-layer routing and protection strategy to optimize IP and optical networking synergies and to forward and protect traffic at the most economical layer.

Conclusion

The cloud ecosystem enables digital transformation and the digital economy. However, it requires a highly interconnected data center infrastructure to enable customers in any location to access cloud applications and services whether located in edge or core data centers. This requires global and regional networks that connect and interconnect data centers in a scalable, secure and reliable way. Moreover, network automation and interactive customer portals are essential to give customers more dynamic access to the cloud and to ensure providers can control operations costs.

By implementing a single, integrated IP/optical network and multi-layer control and automation, data center providers and network operators can implement scalable, flexible and programmable data center interconnection and enable seamless access to applications and services wherever they are located. This significantly enhances the customer experience, reduces operations costs and improves operations efficiency.

Related resources

- Data center interconnection for the cloud era. eBook
- Data center interconnection for the cloud era. Use case
- Achieving efficient IP-optical network automation. Application note.

Abbreviations

AES	Advanced Encryption Standard
AI	artificial intelligence
API	Application Programming Interface
AR	augmented reality
DCI	data center interconnect
ECMP	equal-cost multi-path routing
GMPLS	Generalized Multiprotocol Label Switching
IoT	Internet of Things
IP	Internet Protocol
IX	internet exchange
MPLS	Multiprotocol Label Switching
OSS	operations support system
OTN	optical transport network
PCE	Path Computation Element
RAN	radio access network
RSVP-TE	Resource Reservation Protocol - Traffic Engineering

SDN	software defined network
SLA	Service Level Agreement
SR	segment routing
SRLG	Shared Risk Link Group
VLAN	virtual local access network
VPN	virtual private network
VR	virtual reality
WAN	wide area network

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: (August) CID206944