The journey towards smart roads: Evolving communication networks for safe, on-time travel

White paper

Road operators are working to make smart road systems a reality by embracing digital-era technologies such as vehicle to network to everything (V2N2X), autonomous vehicles (AVs) and AI-assisted Intelligent Transportation Systems (ITS). These technologies promise to help them revolutionize the travel experience and ensure safe, on-time, connected and sustainable journeys for all highway users.

To turn this promise into reality, operators need to make sure their mission-critical IP/MPLS wide area networks (WANs) can keep pace with the connectivity demands of physical and digital highway infrastructure and counter more complex cyber threats. This paper explains how new WAN capabilities and innovative use cases such as segment routing, operational technology (OT) cloud networking, networking for AI workloads, and zero-trust quantum-safe networking can help road operators use their WANs to unlock the full potential of smart roads.

Contents

Introduction

The advent of smart road systems ushers in a transformative era in transportation infrastructure. Road systems, comprising highways, roads, tunnels and bridges, have long served as the arteries of society, connecting urban centers, towns and rural areas. This critical infrastructure facilitates the flow of people, goods and ideas. As society enters the digital age, the road and highway landscape is evolving into an intelligent, connected digital ecosystem, harnessing the power of cutting-edge technologies for safe, on-time, connected and sustainable journeys. The key technologies driving this transformation include vehicle-to-network-to-everything (V2N2X), autonomous vehicles (AVs) and artificial intelligence (AI)-assisted Intelligent Transportation Systems (ITS).

V2N2X

First coined by the 5G Automotive Association (5GAA), V2N2X is a general term for cellular network-based communications paradigms. It leverages cellular networks, particularly 5G, to enable vehicles to communicate with different entities in their immediate vicinity and beyond. For example, vehicles can exchange information with other vehicles, road infrastructure owner/operators (IOOs) and cloud-based services operated by navigation service providers (SPs) and automotive manufacturers (known as original equipment manufacturers, or OEMs). This technology enables advanced applications that enhance safety and traffic efficiency by providing drivers with comprehensive, real-time awareness of the broader traffic environment.

For example, a traffic event information-sharing application allows vehicles, IOOs, SPs and OEMs to exchange information on hazards, school buses, wrong-way drivers and broken-down vehicles. Similarly, a traffic signal information-sharing application shares real-time status of traffic signals with IOO applications (e.g., traffic light controllers) and vehicles. This enables better coordination and optimization of traffic flow, leading to reduced congestion, improved road safety and better fuel efficiency.

Autonomous vehicles

AVs represent a groundbreaking shift in road transportation. They use state-of-the-art technologies to operate without driver intervention. The use of V2N2X technology, particularly through 5G networks, provides AVs with high-speed, low-latency broadband connectivity to communicate with other entities.

Furthermore, AVs use an array of sensors, including LiDAR, cameras, radar and GPS, to perceive their surroundings, make real-time decisions, and navigate complex, fluid environments. AVs can utilize V2N2X connectivity to share this information with other entities in real time, without delay.

As AVs continue to evolve and become more prevalent, they benefit society by making road travel safer, increasing transportation equity and inclusivity and reducing emissions.

AI-assisted C-ITS

The ITS concept can be traced back to the development of early traffic management systems in the 1960s. These systems focused on basic traffic signal control and have since evolved into sophisticated, data-driven traffic management systems that incorporate new capabilities.

Today's ITS, often referred to as Cooperative ITS (C-ITS), represents an evolving, comprehensive approach to traffic management. The systems collect extensive volumes of data from a wide array of sensors and intelligent roadside equipment, process this information and help road operators manage traffic and respond to situational changes ranging from sudden inclement weather to road accidents and traffic congestion. For example, based on historical and real-time roadside data, C-ITS can dynamically implement

variable speed limits and route diversion plans to smooth traffic flow and improve road safety. C-ITS can also adapt traffic light control to dynamically adjust signal timing to shorten wait times.

As roadside sensors generate larger, more diverse datasets, it becomes challenging to process all the data in real time. Consequently, vendors of C-ITS are increasingly embracing AI and machine learning (ML) to enable instantaneous data analysis. By integrating AI with predictive modeling, data visualizations and other decision support systems, AI-assisted C-ITS allows operators to adopt more sophisticated traffic management strategies, maintain infrastructure predictively and respond to changing road conditions.

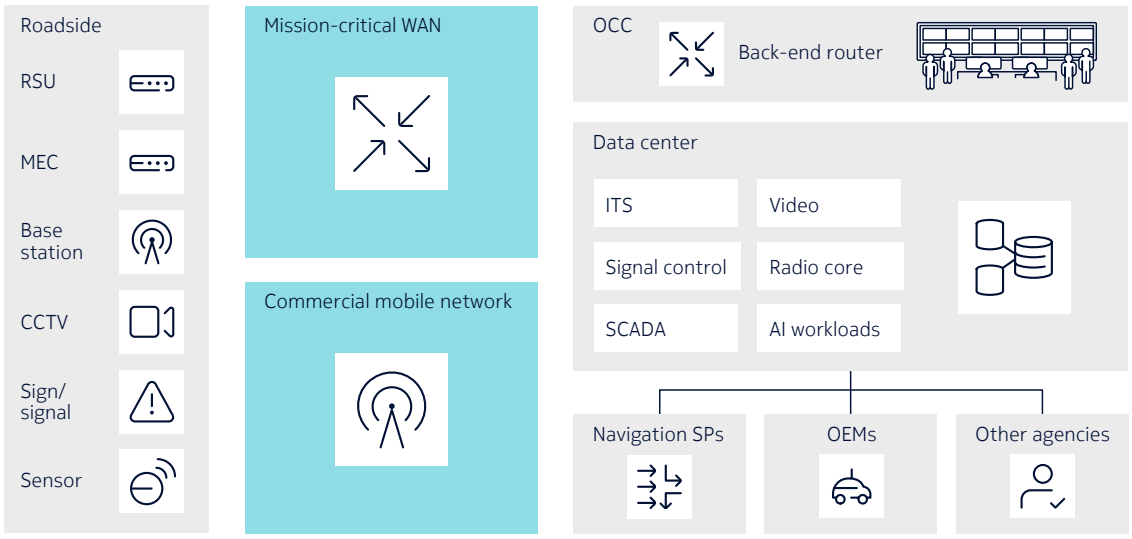
Smart roads are going digital

The smart highway system harnesses these latest technologies to enable software-centric and data-driven operations. It comprises extensive physical infrastructure, including roads, intelligent roadside equipment and sensors across the highway system where data is generated and exchanged, as well as digital infrastructure where data is consumed by central software systems.

Adoption of V2N2X ushers in the use of multi-access edge computing (MEC), which extends the digital infrastructure footprint from central operations centers and data centers toward the roadside. At the same time, V2N2X introduces the need to extend digital connections beyond the operator's own domain. It disseminates data through an information-sharing domain to interconnected SPs, OEMs and other government agencies.

The mission-critical wide area network (WAN) infrastructure plays a central role in the smart highway by connecting all physical and digital infrastructure elements. Figure 1 shows a high-level system view of a smart highway and its associated ecosystems.

Figure 1. Smart highway system architecture



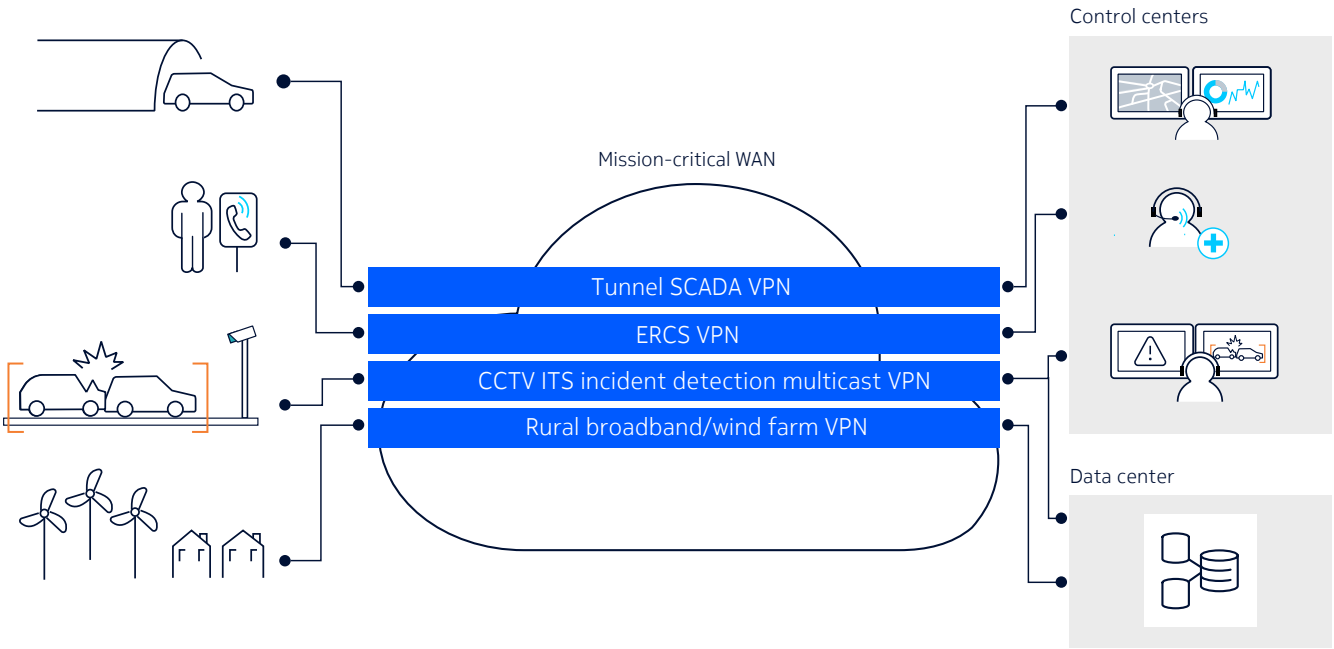
RSU: roadside unit
ITS: Intelligent Transportation Systems
MEC: multi-access edge computer
OEM: original equipment manufacturer
SP: service provider

Mission-critical WAN core essentials

The WAN is a multilayer network that features IP/MPLS over dark fiber, packet optical transport and packet microwave transmission. It is built with the following essential mission-critical networking capabilities:

- **Multiservice capability**: The WAN needs to support a mix of new IP-based applications such as ITS and CCTV, as well as older SCADA and emergency roadside call system (ERCS) applications that still operate with legacy serial and four-wire analog interfaces respectively (Figure 2). It also needs to support flexible layer 2 and layer 3 network services with highly scalable connections for a plethora of roadside equipment and sensor endpoints. Additionally, IP multicast capabilities are essential for efficient point-to-multipoint distribution for applications such as CCTV. If opportunities arise, these capabilities can be used to bring connectivity to rural populations and renewable energy facilities.

Figure 2. A multiservice WAN serving numerous highway applications



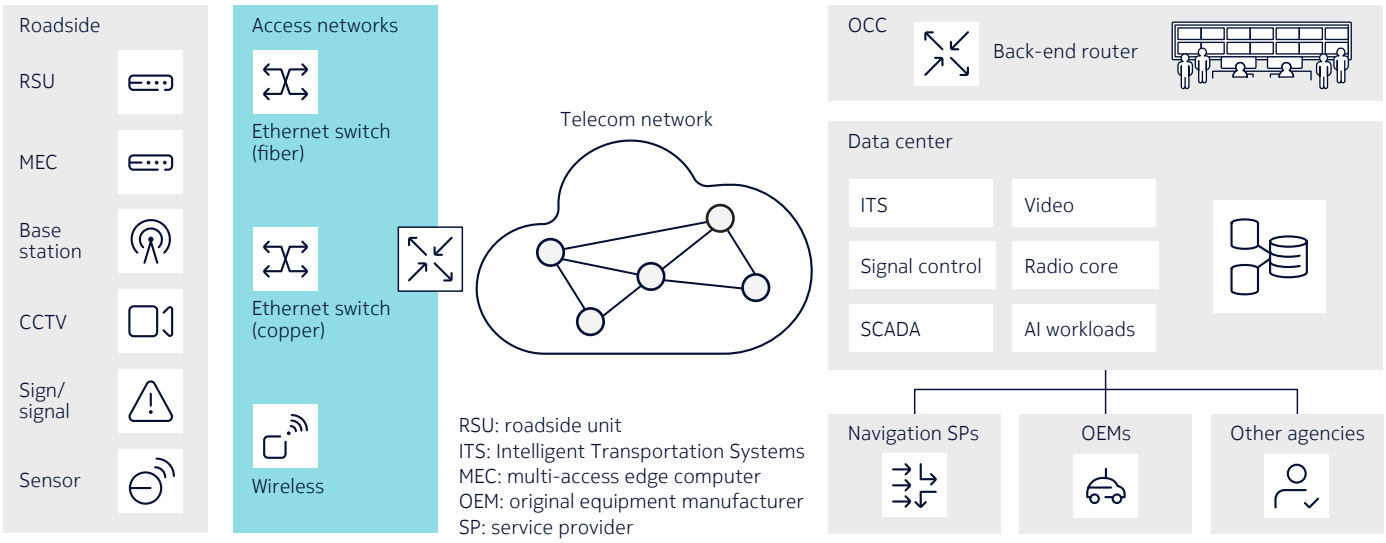
- **Strong resiliency**: A WAN outage would cause the operator to lose visibility of highway conditions and could have grave safety consequences. The WAN can leverage the full capabilities of IP/MPLS multilayer/multifault redundancy protection to ensure high network availability and withstand extreme weather events and deliberate fiber cuts.
- **Deterministic quality of service (QoS)**: The WAN needs robust and flexible QoS capabilities to constantly meet the network performance requirements of a diverse set of applications, including voice and real-time ITS applications that have strict latency and lossless networking requirements even when the network is congested (Table 1).

Table 1. A reference QoS profile for smart highway applications

Applications	Latency	Bandwidth	Reliability	Criticality
ERCS	Low	Low	High	High
CCTV	Medium	High	Medium	Medium
DMS	High	Low	High	High
Toll collection	Medium	Low	High	Low
Weather sensor	High	Low	Low	Low

- **Diverse access media**: A smart highway mandates ubiquitous connectivity throughout the expansive highway infrastructure. The WAN connects with various access domains from copper cable wire to optical fiber to commercial and private wireless links (Figure 3).

Figure 3. Diverse access media to provide ubiquitous connectivity everywhere



Evolving the WAN for advanced smart highway communications

Road operators need to evolve their WANs to support the growing demands of smart highway innovations and counter increasingly sophisticated cyber threats. As smart highway applications incorporate more advanced technologies, the underlying network infrastructure needs to adapt to provide enhanced capabilities, robust security and seamless connectivity into data centers. This WAN evolution is crucial to enable the full potential of smart highways.

Harnessing the full power of segment routing in IP/MPLS WAN

As they deploy more intelligent roadside equipment such as roadside units, sensors, wireless access points and base stations, road operators continue to expand their networks. One proven strategy is for operators to extend the IP/MPLS network with segment routing to increase service scalability and add network capabilities to support new digital innovations.

Segment routing offers an approach for distributing labels to routers in IP/MPLS networks. It uses routing protocols (OSPF or IS-IS) that already run in the network and adds segment routing extensions to distribute labels, also known as segment IDs (SIDs). By simplifying label management and distribution, segment routing enables the network to scale massively to support future growth. IP/MPLS networks can seamlessly migrate to Segment Routing over MPLS (SR-MPLS) without impacting the existing network hardware and routing design. The migration simplifies path establishment and network scaling and enables better control of data traffic.

When combined with a Path Computation Engine function in the transport network manager, the IP/MPLS transport network now provides enhanced capabilities for network path computation and traffic engineering. It enables more effective load sharing with high data flow granularity by allowing fine-grained control over packet forwarding paths. IP/MPLS also supports end-to-end traffic engineering for paths that span multiple routing areas, a common design practice in large-scale networks.

To accommodate the ever-increasing number of roadside devices with higher physical network port fan-out, operators should consider an aggregating industrial Ethernet network layer as an integrated, essential part of the WAN. This layer can enable end-to-end network service while maintaining robust redundancy and security measures.

OT cloud networking

Many road operators now deploy critical smart highway applications such as ITS and data analytics in a set of segregated on-premises servers that operate in a cloud environment hosted within their data center. This environment is known as operational technology (OT) cloud.

OT cloud is a key component of the highway digital infrastructure. It serves as a platform for:

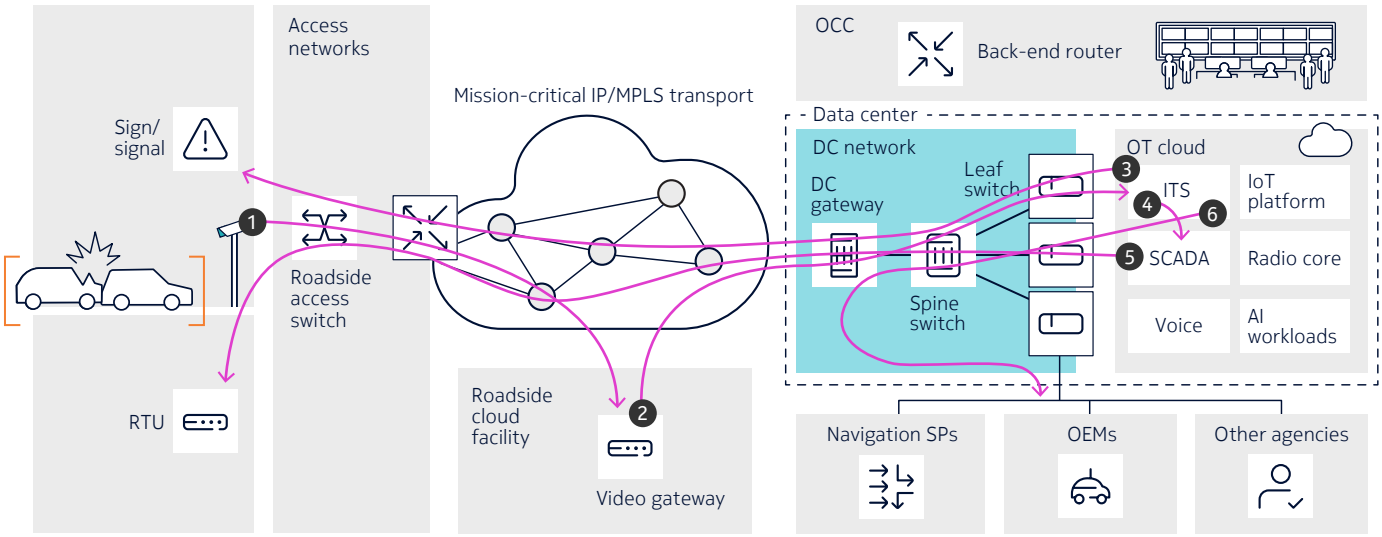
- Enhancing application performance and agility, which enables operators to adapt to a changing operational environment without compromising the availability and performance of critical applications
- Embracing emerging and innovative technologies such as AI and ML.

Because applications hosted in OT cloud continually communicate with roadside assets and sensors, the data center network is now part of the mission-critical connectivity infrastructure. The data center network needs to interwork seamlessly with the IP/MPLS network to provide end-to-end mission-critical connectivity for highway applications.

OT cloud networking is a response to this new requirement. It extends mission-critical connections from the IP/MPLS transport network into the data center network, which is composed of the data center gateway and data center fabric. The data center fabric is made up of leaf switches that connect to servers and spine switches that serve as the aggregation layer. The data center gateway serves as the interworking gateway between the fabric and the IP/MPLS WAN.

Through the gateway, the backbone can harness Ethernet Virtual Private Network (EVPN) service and Border Gateway Protocol (BGP) routing in IP/MPLS transport and data center network domains to seamlessly interwork with the fabric at the service and IP layers. This interworking gives the network the agility required to support dynamic OT cloud networking that connects the radio access network (RAN) to the core domains. Figure 4 illustrates an OT cloud networking deployment in support of an incident detection and response use case.

Figure 4. OT cloud networking communications supporting smart highway applications



1. CCTV cameras continuously stream video information to video gateway inside a roadside cloud facility for road surveillance to detect incidents.
2. The video gateway will alert ITS inside the data center when an incident is detected.
3. The ITS signal control subsystem programs dynamic signs to display messages that lower the speed limit and warn drivers that there is an accident ahead.
4. The ITS also notifies the SCADA system if the accident occurs in a tunnel.
5. The SCADA server commands the SCADA remote terminal unit (RTU) to activate tunnel emergency systems such as ventilation systems.
6. The ITS shares road accident information with other third parties in the ecosystem to notify road users.

Networking for AI workloads

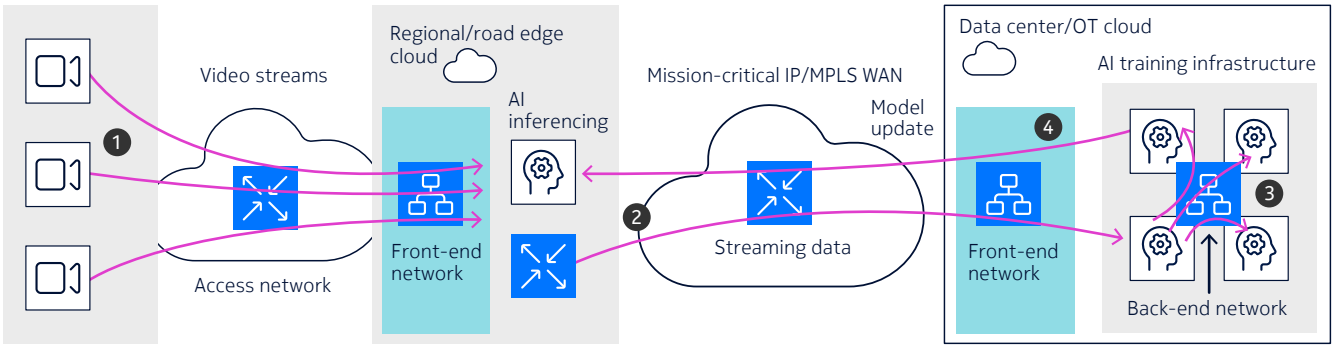
AI, which encompasses branches ranging from deep learning and computer vision to generative AI and artificial general intelligence (AGI), offers immense potential to enhance highway operations. It can help road operators make faster, data-driven decisions that improve safety, efficiency, equity, resiliency and sustainability for road transport.

AI deployment in highway systems involves two key phases: training and inference. The process begins with data collection and training, where operational data from sensors, cameras and historical sources is gathered and processed using on-premises graphics processing unit (GPU)-accelerated servers or AI platform services from cloud providers (GPU-as-a service or GPUaaS). A high-performance network is essential for efficient data processing and model development.

Once the models are trained, they are strategically deployed across central, regional or roadside facilities to perform real-time inference that enables predictions and anomaly detection. This deployment relies on robust front-end networking to ensure seamless data flow and rapid response times. Road operators can facilitate continuous learning by regularly updating the models with new data, which enhances their accuracy over time. This comprehensive approach ensures effective AI implementation and improves highway operations by enabling data-driven decision-making.

The use of AI in highway systems requires an extensive network infrastructure that connects OT cloud to the roadside edge cloud and to sensors across the highway system. Figure 5 provides a high-level view of the data flow for an AI workload for traffic accident detection.

Figure 5. Networking for highway AI applications



1. CCTV camera streams multicast video data over the access network and the front-end network to the roadside edge cloud, where an AI application such as a video gateway detects accidents and other anomalies.
2. The CCTV camera also sends the video data to the AI training infrastructure over OT cloud, the front-end network and the WAN.
3. Within the AI training infrastructure in the OT cloud, data is forwarded from one GPU to another over the back-end network.
4. The AI training infrastructure sends model updates to the AI workloads at the edge from time to time.

AI workloads are computationally intensive and focus on neural network training and inference. Training involves large-scale matrix multiplications across GPU clusters using parallel processing.[1] Inference, while less intensive, requires rapid data processing for real-time tasks such as detection and prediction.

The distributed nature of these operations, combined with the need for real-time data exchange, places significant demands on the communication infrastructure. Efficient data transport is crucial for maintaining AI system performance and scalability. Even minor communication bottlenecks can significantly impact training efficiency and inferencing response times. This underscores the importance of having a robust, high-performance network to support AI applications in highway operations. The network infrastructure must enable reliable real-time data delivery for interconnecting GPUs used for training and inferencing.

Utmost resiliency

The network plays a pivotal role in reliably delivering numerous high-resolution video streams to video analytics applications in the edge cloud for inferencing. It also delivers these streams to the central AI training infrastructure for continuous learning. Any network outage could compromise the timely detection of anomalies or inclement road conditions and hinder AI inferencing performance. It would also directly jeopardize safety and efficiency. Moreover, a network failure can disrupt the training cycle. To mitigate such risks, the network needs to have rich path diversity and support advanced redundancy protection capabilities ranging from MPLS fast-reroute to equal cost multipath.

1 For traffic accident detection tasks, matrices represent sequences of video frames, with each element corresponding to pixel values. GPUs perform rapid parallel calculations on these image matrices to identify accidents or anomalies.

High network capacity

AI training involves ingesting vast datasets. This task can extend over days or weeks for initial training. GPUs are precious resources, so minimizing job completion time (JCT) becomes crucial. To achieve this goal, the network infrastructure, particularly the back-end network, needs to support high-speed interfaces ranging from 100GE to 800GE to increase network capacity. Efficient handling of the massive data flows between compute nodes optimizes parallel processing and ultimately reduces JCT.

Congestion-free, lossless networking

Network congestion leads to packet loss and impacts AI workload performance. In the WAN, packet loss can hinder edge AI inferencing workloads and make it harder for the system to detect anomalies.

Fortunately, there are effective network strategies for tackling congestion. With a network services platform, operators can intelligently provision connections across the network based on the service intent, including bandwidth and class of service. Network analytics and QoS capabilities such as intelligent packet discard and buffer management can also play a crucial role in prioritizing AI application data.

The situation in the data center, where AI workloads are distributed across a GPU cluster for parallel processing, is more challenging. Since each GPU handles only a portion of the data or model, all the GPUs in the cluster need to synchronize their computations in real time. This synchronization is crucial for neural network model training, where the output of one layer in the neural network becomes the input for the next, potentially spanning multiple racks of GPU in the data center. As a result, there are frequent instantaneous data bursts in the fabric as each GPU computes, processes and sends out data to the next layer. These data bursts can cause congestion and packet loss in the fabric.

If packet loss occurs, the receiving GPU stops processing and waits until the sending GPU retransmits the discarded packets. This idling time delays other GPUs from advancing the process pipeline, which increases JCT.

It is critical to eliminate the need for retransmissions in the data center network. If congestion does occur, the back-end network can use advanced congestion control and notification mechanisms such as Priority-based Flow Control (PFC), Explicit Congestion Notification (ECN) and Data Center Quantized Congestion Notification (DCQCN). In addition, advanced hashing techniques can help evenly distribute traffic across the set of switch uplink interfaces.

Low overall and tail latency

Another measure for reducing JCT is to ensure minimal latency in the back-end network when switching data among GPUs to maximize GPU utilization. Network latency is caused by transmission and switching in the network node. High-speed interfaces can reduce transmission latency by speeding up the transmission of Ethernet frames.

Tail latency and switching latency are other factors to consider. Tail latency refers to the high end of the latency distribution, typically measured at the 99th percentile or higher. It represents the worst-case latency experienced by a small portion of data transmissions. As explained earlier, GPUs in the cluster often operate in parallel, and high latency experienced by just one GPU can delay other GPUs by forcing them to wait. This is particularly critical at the training stage, where intensive matrix multiplications are performed, and the completion time is determined by the last GPU to finish multiplying its assigned matrix elements. Consequently, high tail latency can hinder the progress of the entire training process. A robust, high-throughput and low-tail latency network infrastructure is essential for ensuring reliable real-time data delivery to facilitate seamless inter-GPU communication for training and inferencing.

A non-blocking network topology is essential for reducing switching latency. GPUs used in AI training often generate large data flows. It is not uncommon to connect the edge of the fabric, composed of leaf switches, to each GPU with a 400 GE link. A non-blocking topology needs equally high bandwidth in the fabric core, composed of spine switches, to avoid blocking.[2] A non-blocking topology is an architectural design that provides sufficient bandwidth and eliminates oversubscription. Coupled with advanced hashing algorithms to distribute traffic evenly, it ensures any incoming data can be switched to its intended destination without contention or queuing, regardless of the traffic patterns in the network.

Seamless WAN/AI infrastructure integration

AI deployments in highway systems require connections that extend from the central OT cloud to the entire highway infrastructure. A data center gateway bridges the cloud network and the WAN at the transport, IP and service layers. It plays a pivotal interworking role by providing seamless end-to-end connectivity for AI workloads.

Zero-trust WAN security

Highway systems are high-value targets for malicious actors in cyberspace. As highway operations become increasingly digitalized, the attack surface of highway infrastructure expands significantly. Protecting the confidentiality, integrity and availability of application data as it traverses the transport network is paramount for ensuring safe and efficient digital highway operations.

An impregnable cyber defense requires a zero-trust approach combined with a multilayer defense-in-depth framework that extends across the infrastructure, networks and applications. Road operators can turn the WAN into the first line of cyber defense by implementing a comprehensive suite of security tools, including network segmentation through IP/MPLS services, IP and Media Access Control (MAC) filtering and role-based network management.

However, as bad actors gain access to more sophisticated resources and evolve their attack methods, the security posture needs to adapt to address current and future threats. With the arrival of cryptographically relevant quantum computers (CRQCs) looming, many current network security measures will soon become vulnerable. To ensure an effective defence against current attacks and emerging quantum threat, the transport network security posture needs to be strengthened with quantum-safe encryption.

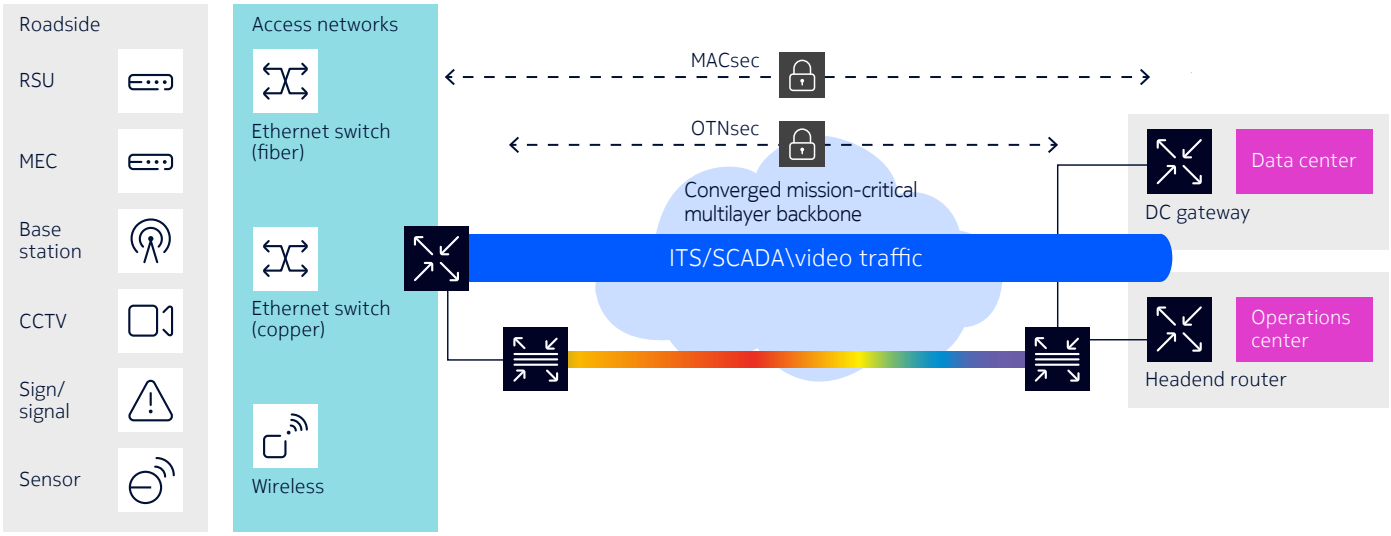
Traditional encryption methods have been effective against various types of attacks, including eavesdropping, man-in-the-middle (MITM) and denial of service (DoS). However, the rapid advancement of quantum computing and quantum algorithms is poised to upend the cybersecurity landscape. For example, a bad actor running Shor's algorithm on a quantum computer could break the protection provided by asymmetric key encryption schemes such as Diffie-Hellman, making the road system vulnerable to eavesdropping.

The commercial value of data may not seem significant, but once bad actors capture and analyze the data, they can begin launching targeted MITM and DoS attacks on critical systems such as ITS. These attacks have the potential to disrupt road operations and jeopardize safety.

It may take years for standards organizations to finalize post-quantum cryptography algorithms for the application layer, but road operators can protect their infrastructure against quantum attacks today. A multilayer defense-in-depth approach that uses Layer 1 OTNsec and Layer 2 MACsec with symmetric AES-256 encryption can provide quantum-safe protection. This approach can thwart quantum attacks by ensuring security across the infrastructure, from data centers and operation centers to roadside equipment (Figure 6).

2 Blocking occurs when two Ethernet frames arrive at the egress port at the same time and contend for the next transmission slot.

Figure 6. Multilayer defense-in-depth encryption in the transport backbone



Conclusion

The journey towards smart roads promises to advance and transform transportation infrastructure. By embracing cutting-edge technologies such as V2N2X, AV and AI-assisted C-ITS, road operators will revolutionize the travel experience.

At the heart of this journey lies a resilient, high-capacity and secure mission-critical IP/MPLS WAN. As the WAN evolves to support OT cloud, AI integration and quantum-safe communications, it will need to extend robust and secure connectivity from intelligent equipment and sensors at the roadside to smart road applications and emerging AI workloads in OT cloud and human workers in operations centers.

Nokia addresses this need with a broad product portfolio that spans IP/MPLS, data center fabric, packet optical, microwave, 5G, LTE, security, IoT and analytics. We complement this portfolio with a full suite of professional services, including network audit, design and engineering practices. With this combination of proven technology and expertise, we offer unique capability and flexibility to help road operators transform their networks to support smart highways for the digital future.

To learn more about Nokia smart road and IP/MPLS solutions, visit our Highways solution and IP Networks web pages.

Abbreviations

5GAA	5G Automotive Association
AES	Advanced Encryption Standard
AGI	artificial general intelligence
AI	artificial intelligence
AV	autonomous vehicle
BGP	Border Gateway Protocol
C-ITS	Cooperative Intelligent Transportation Systems
CCTV	closed circuit television
CRQC	cryptographically relevant quantum computers
DCQCN	Data Center Quantized Congestion Notification
DMS	Dynamic Message Signs
DoS	denial of service
ECN	Explicit Congestion Notification
ERCS	emergency roadside call system
EVPN	Ethernet Virtual Private Network
GPS	Global Positioning System
GPU	graphics processing unit
GPUaaS	GPU as a service
IOO	infrastructure owner/operator
IP	Internet Protocol
IS-IS	Intermediate System to Intermediate System
ITS	Intelligent Transportation Systems
JCT	job completion time
LiDAR	Light Detection and Ranging
MAC	Media Access Control
MACsec	Media Access Control Security
MEC	multi-access edge computing
MITM	man in the middle
ML	machine learning
MPLS	Multiprotocol Label Switching
OCC	operations control center

OEM	original equipment manufacturers
OSPF	Open Shortest Path First
OT	operational technology
OTNsec	Optical Transport Network Security
PFC	priority-based flow control
QoS	quality of service
RAN	radio access network
RSU	roadside unit
SCADA	Supervisory Control and Data Acquisition
SID	segment identifier
SP	service provider
SR-MPLS	Segment Routing over MPLS
V2N2X	vehicle to network to everything
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 is a registered trademark of Nokia Corporation. Other product and company names mentioned herein may be trademarks or trade names of their respective owners.

© 2024 Nokia

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

Document code: (November) CID214347