Deloitte.

Security and compliance in 5G and AI-powered edge networks

February 2024

At Mobile World Congress 2024, Deloitte demonstrates enhancing network security and facilitating compliance with industry regulations in connected edge networks, aimed to help enterprises throughout their connected edge modernization journey.

Contents

Abstract

Use cases enabled by connected edge networks require cybersecurity measures to protect applications, connected devices, machine learning models, wired and wireless access, and data, while delivering full network visibility to detect and respond to threats in real time.

The evolution of advanced connectivity, as seen, inter alia, in 5G and AI-powered connected edge networks, is shaping leading practices for cybersecurity and regulatory compliance. The adoption of edge use cases aims to enhance consumer experience through ease, speed, and convenience. This paper explores different cybersecurity strategies, describes the threat landscape, and discusses vectors that are shaping the security architecture for edge networks. To help organizations navigate the evolving regulatory landscape, cybersecurity standards, and industry-specific frameworks, Deloitte is bringing forward leading practices to help comply with the Payment Card Industry Data Security Standard (PCI DSS) and consumer privacy acts and to adopt zero trust architectures. Implementing security controls at the edge requires proper network segmentation, Internet of Things (IoT) security, secure wired and wireless access, application security, and advanced network security. Additionally, secure software development lifecycle, supply chain, and AI workloads are discussed to highlight the proper network configuration principles that can be implemented to help mitigate security risks and address security threats.

Connected Edge Use Cases

Edge computing and AI technologies in industries such as retail aim to unlock value through effective, personalized customer experiences and enhanced operations for employees.

It is expected that, by 2026, 90 percent of the top two thousand retailers will utilize edge computing to harness data in stores [1]. Deloitte's connected edge initiative addresses the pressing issues across the retail industry, including the goal of enhancing consumer experience through ease, speed, and convenience. Alongside changing consumer behaviors, the competitive market is compelling store management to incorporate automation and equip workers with new skills, capabilities, and opportunities to excel in their roles. To cater to demands and adapt to changes, leading retailers are moving towards a service approach that is convenient, efficient, and customer focused. This can be observed in the specific trends outlined in Deloitte's 2023 Retail Industry Outlook [2], which include (1) making investments in camera analytics and sensors to correctly detect and reduce the likelihood of theft and spoilage, (2) cultivating a personal level of customer engagement by customizing products and services based on individual preferences, (3) streamlining the customer checkout process to reduce wait times and queueing, and (4) automating the prioritization and allocation of tasks in real time

Deloitte's approach to connected edge, shown in Figure 1, is built to achieve such near real-time operations through six highly innovative use cases including shrink reduction, hyper-personalization, omni-channel inventory management, store and merchandising optimization, workforce management automation, and checkout transformation.

Insights derived from various data sources, such as cameras and sensors, allow businesses to move from a reactive stance to delays, towards a more predictive, proactive approach. The data is securely processed by AI models for real-time analysis and insights, thereby enabling the workforce to perform their job more effectively.

Deloitte's connected edge initiative pulls together a diverse ecosystem of more than 15 retail technology vendors to provide an interoperable and integrated outcome. The integrated platform runs diverse applications to power the following:

- State-of-the-art edge computing
- Advanced 5G and Wi-Fi connectivity
- AI acceleration with computer vision and multiple sensors
- On-site data processing

By leveraging sensors and cameras to collect data from retail shelves, in-store processing allows instantaneous analyses and insights, thereby enhancing the efficiency, productivity, and effectiveness of the workforce.

Deloitte's connected edge initiative is an example of a 5G, AI-powered edge network that is designed with a sharp focus on security and compliance. Securing the edge infrastructure, shown in Figure 1, and its AI-powered applications while complying with industry standards and regulations is discussed in the following sections.

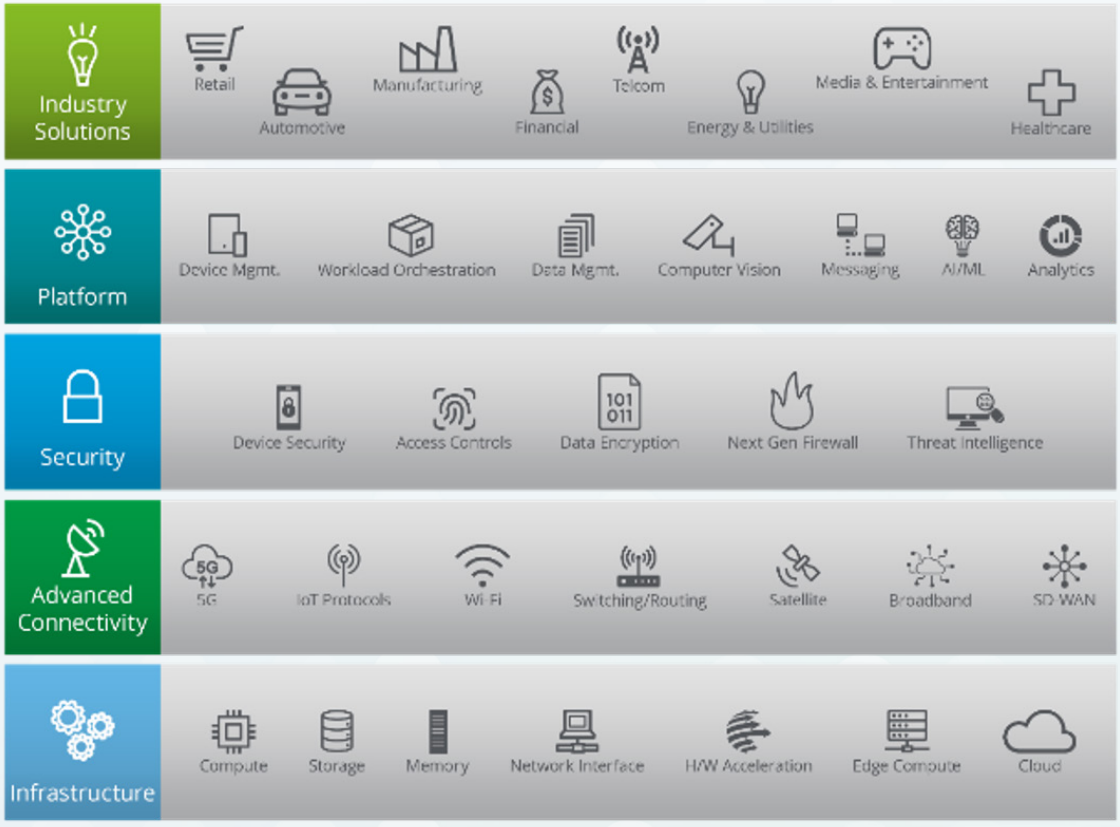

Figure 1. Deloitte's Connected Edge architecture

Security Governance

Harnessing the next generation of wireless and AI-powered edge technologies requires enterprises to comply with new regulations, industry-specific guidance, and leading cybersecurity standards.

Leveraging guidelines from leading connectivity standards bodies, national cybersecurity groups, as well as industry-specific guidelines requires a harmonized cost-effective security and privacy framework to guide enterprises through their connected edge transformation.

Securing edge networks requires a mesh of core principles and leading practices starting with governance, Service Level Agreements (SLAs) within multi-tenant environments, and the adoption of a Zero Trust (ZT) framework that ties different standards and regulations to industry-specific controls. The complex ecosystem of 5G and connected edge networks requires a hybrid governance model that maps security boundaries, responsibilities, and accountability for telco, multi-cloud, edge, AI/ML and IoT vendors. Certain mission-critical use cases require advanced connectivity such as a private 5G network deployment, making SLAs and contractual terms related to network security and performance critically important for Chief Information Security Officers (CISOs) [3]. Systems running on containerized or virtualized deployments present a myriad of intersecting responsibilities between teams such as access control, event logging, inspection, configuration management, supply chain risk management, and Personally Identifiable Information (PII) processing, and data transparency.

Cybersecurity standards and leading practices related to 5G and connected edge security continue

their evolution fueled by different industry-specific, national, and international bodies. To list a few, the next generation of AI-powered devices, wireless communication, and edge technologies require compliance with advanced standards and regulations established by bodies such as the Cybersecurity and Infrastructure Security Agency (CISA), National Institute of Technology (NIST), National Security Agency (NSA), International Organization for Standardization (ISO), Payment Card Industry Data Security Standard (PCI DSS), Global System for Mobile Communications (GSMA), and 3rd Generation Partnership Project (3GPP).

PCI DSS is a set of security standards designed to determine if companies that accept, process, store or transmit credit card information maintain a secure environment. PCI DSS applies to organizations, regardless of their size or number of transactions, that accept, transmit, or store cardholder data, according to the latest version of PCI DSS 4.0, released on March 31, 2022 [4]. Businesses that manage payment card information are required to achieve and maintain PCI compliance to guard the security of sensitive data and reduce the risk of data breaches. Compliance is an ongoing process and not a one-time event. Hence, businesses are required to regularly assess and inspect their compliance status. While the standard covers areas such as network security, access control, data encryption, regular testing, and monitoring, two security domains stand out in a connected edge use case. The first domain is identity and access management (IAM) for edge nodes and clusters storing, processing, or transferring payment information. The second domain touches data in its three forms - at rest, in transit, and during processing. In addition, physical assessments may be required to determine that no PII could be accidentally captured by existing or newly installed cameras. Similarly, in containerized deployment, such as Kubernetes, designing PCI compliant architectures may require virtual network segmentation, namespace separation, and the provisioning of various credentials for system administrators. This can help isolate pods handling PII requests from image processing pods used for AI/ML models to power edge use cases.

Mobile network operators, on the other hand, plan to offer dedicated 5G end-to-end network slices to enable in-store applications which intersect with the latest guidance published by NSA and CISA on security considerations for design, deployment, and maintenance of 5G network slicing [5]. From the enterprise perspective, assessing the service quality and security policy rules of each network Slice/Service Type (SST) gives enterprises the requisite reliability and confidence to stay in compliance when introducing new, 5G-based edge use cases. Securing 5G network slicing is discussed in the next section.

Complying with regulations also includes privacy legislations such as the General Data Protection Regulation (GDPR) in the European union or California Consumer Privacy Act (CCPA), particularly when video surveillance and analytics are driven from cameras monitoring consumer activity or purchase behavior. Obtaining explicit consent from consumers while ensuring that their privacy is respected constitutes a major requirement of compliance with privacy laws. Such compliance requires a balanced approach that adheres to privacy laws while maintaining cybersecurity requirements for data including confidentiality, integrity, and availability.

Connected Edge Security

Securing computer vision use cases in decentralized edge networks powered by 5G or Wi-Fi connectivity requires real-time traffic visibility, advanced threat correlation, and automated security enforcement.

Edge computing takes data processing and application execution closer to the end user. Instead of using a centralized cloud service, the workload is decentralized and dealt with at the 'edge' of an organization's network. In our retail store example, the data storage and processing, coming from the different connected devices, takes place at the store's local network, with dedicated Processing Units (CPUs) and Graphics Processing Units (GPUs), networking, and storage requirements. This provides lower latency and faster processing capabilities for AI models. As with other use cases, securing the retail store at the edge extends beyond users and devices to the cloud and application fabric by implementing cloud-native security for workloads and integrating next-generation firewalls to help protect against known and unknown inbound and outbound threats, prevent data exfiltration, protect individual devices and prevent lateral movement across applications and within the network [6].

The complexity of the example presented makes the implementation of a Zero Trust Architecture (ZTA) a challenging task that requires a model that is built upon strong foundational capabilities across five fundamental pillars: identities, workloads, data, networks, and devices. Telemetry & analytics, and automation and orchestration intersect with the five pillars to monitor, analyze, and respond proactively to threats. Due to the need to enforce access

requirements in edge locations, cloud resources should be designed with the proper access control upon provisioning and teams should use Infrastructure as Code (IAC) methods with tightened access policies to prevent compromises on security. The benefits of adopting ZTA and IAC include:

- Minimizing the attack surface, containing, and reducing manual deployments that lead to human errors.
- Enforcing least privilege access to resources along with continuous monitoring to identify, contain, and prevent cyber-attacks [7].
- Managing and monitoring least privilege access including provisioning and revoking access to the authorized list of users of the connected edge and maintaining separate credentials for different PCI-tagged applications.
- Minimizing the number of admins within the store and providing floating provision of least privilege access (only as and when needed) reducing users exercising admin rights for longer durations.

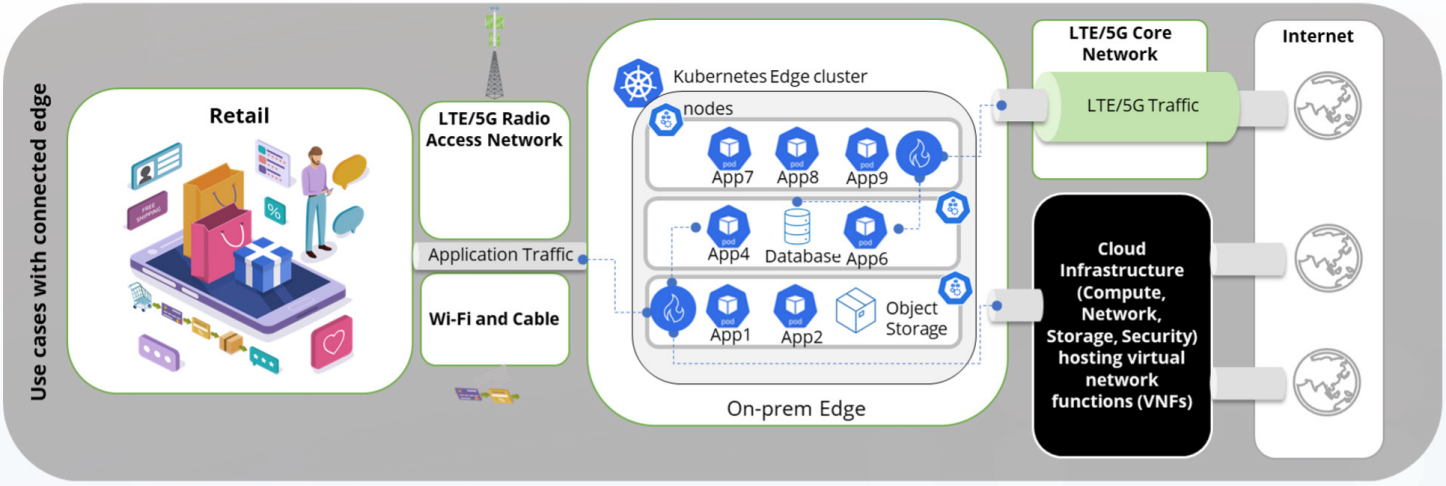

Figure 2. Containerized edge architecture

The retail store use-case typically uses a wide range of devices such as cameras that are deployed at security, billing, and inventory checkpoints (on shelves) for replenishing items, QR code scanners, and various temperature and humidity sensors. The store infrastructure, highlighted in Figure 2, may also include on-premises servers, Point of Sale (POS) systems, and a modern edge infrastructure running Kubernetes clusters with respective data pipelines or other connected services, the networking devices, and internet connections. This can expand to a myriad of combinations of these devices and applications as the business scales. One common requirement for most of these devices is continuous software updating and patching, which represents a major risk if secure software and hardware supply chain is not adopted. Having next-generation firewalls and zero-trust network segmentation agents monitor and control device access to and from the internet and inter-device connectivity may help reduce code injection and command and control attacks. IoT asset profiling using advanced AI/ML models helps in preventing known vulnerabilities of high-risk ports, and unencrypted communications for the connected devices.

From a network perspective, there are various parts to secure, especially in a cloud-native, Kubernetes cluster. The complexity of network security arises from multitude of threat vectors including North-South and East-West traffic, intra-pod and inter-pod communication, IoT and camera's communication with the cluster, and the cluster's external communication back to the cloud or enterprise network. Implementing virtual and

physical network segmentation between assets and applications including namespace isolation is one of the major design principles to adopt when architecting edge clusters to prevent lateral movement and reduce the blast radius of attacks.

In addition, integrating container-based next-generation firewalls helps (1) protect against known and unknown inbound and outbound threats, (2) prevent data exfiltration, and (3) stop lateral movement across applications and within the network. Real-time traffic visibility and IoT asset profiling provide means to enrich network security policies and zero trust controls that can help remediate vulnerabilities including 0-day, enforce web and DNS protection, and help prevent command and control.

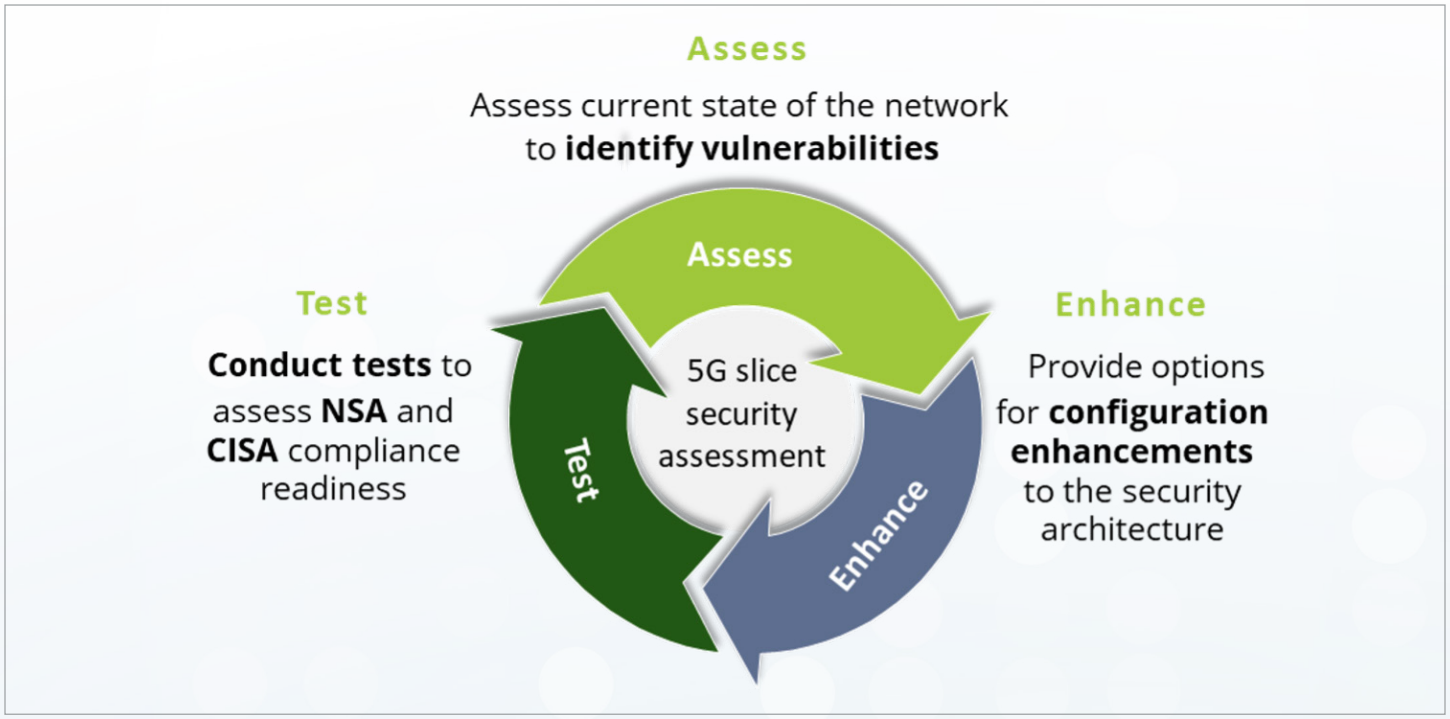

Figure 3. Deloitte's 5G security assessment

5G network slicing is an exclusive feature of 5G networks that may be used in a connected edge use case. Deloitte has developed a three-step approach to assess network slice security - Assess, Enhance, and Test, illustrated in Figure 3 - comprised of over 150 security controls that accelerate and streamline the approach to 5G network and slice security[8]. Inspired on the work developed by National Security Agency and Cybersecurity and Infrastructure Security Agency, Deloitte provides an illustrative non-exhaustive checklist of leading practices that address cybersecurity issues across the different network planes, such as user equipment, data networking and virtualization. The on-prem edge infrastructure often uses Application Programming Interfaces (APIs) to enable IoT applications to connect with different applications and transfer data between workloads, databases, data lakes, and other storage technologies. These APIs are of extreme importance and should be assessed for the proper usage of authentication, tokens, encryption, and secure communication to prevent cyber-attacks. Having full visibility into the network traffic by obtaining real time insight of devices, endpoints, and pod communication allows close monitoring and prevention of malicious activity and abnormal behaviors..

Applications and AI Security

Securing applications and AI workloads requires continuous risk assessments of least privilege access to applications and data integrity checks for AI models.

Content inspection and control over edge applications should be put in place to determine that data is not accessed wrongly and to enhance the detection of malware. Software development should follow a secure integration and development (CI/CD) development pipeline, enabling an ongoing adaptive strategy, loading the network with better features to protect against old and new attacks.

The use of AI is becoming the norm across different areas of application and a fundamental part of modern edge transformation initiative[9]. It can provide huge business advantages but, unsurprisingly, it also brings its own set of vulnerabilities and security concerns that need to be addressed. Typically, AI engines work based on a model that feeds on data and requires extra training data to fine tune what the outputs should be when new input data comes.

It is, therefore, critical to make sure that the training data is not poisoned and that the training stage is not compromised by injection attacks from malicious parties. Furthermore, from an input perspective, it is also

required to prevent attacks that can trick AI engines by perturbing the model with misclassified data. From the output point of view, it is required to assess that the AI engine outcome cannot be retrieved by an adversary and that the output data is not released to users that should not have access to it (need-to-know basis). It is relevant, from a design perspective, to limit the uses and specify the purpose of the AI engine. This avoids using unnecessary data for model training and makes sure that sensitive information is used in an appropriate way.

On top of that, AI development workflow should follow the conventional guidelines and leading practices used for traditional software. This helps prevent attacks such as backdoors, credential theft, use of non-intended models, and contributes for a CI/CD pipeline. Checking the integrity of the data used (for training or just as model input) as well as its access permissions and scope of use is equally important for an AI-secure and compliant environment.

Conclusion

Connected edge computing and advanced connectivity are enabling effective, personalized customer experiences and enhanced operations for various industries. Securing connected edge use cases that are powered by AI workloads requires a broad framework that adopts zero-trust principles and helps organizations comply with regulations, privacy acts, industry-specific guidance, and leading cybersecurity standards. Proper enforcement of cybersecurity measures should encompass applications, connected devices, machine learning models, wired and wireless access, and data, while delivering full network visibility to detect and respond to threats in real time. In addition, security for AI workloads requires continuous risk assessments for least privilege access to ML models and rigorous data integrity checks for training, testing, and real-time data.

Authors

Frederico Macias
Partner
Deloitte Portugal Services Limited
fremacias@deloitte.pt

Ally Adnan
Managing Director
Deloitte & Touche LLP
allyadnan@deloitte.com

Shehadi Dayekh, Ph. D.
Specialist Leader
Deloitte & Touche LLP
sdayekh@deloitte.com

References

01. Deloitte, "Be out front - EDGE.AI accelerates store innovation," 2023.
02. Deloitte, "2023 Retail Industry Outlook," 2023.
03. Deloitte, "3 Critical Elements of Strong 5G Service Level Agreements," The Wall Street Journal, 2023.
04. PCI Security Standards Council, LLC, "Payment Card Industry Data Security Standard: Requirements and Testing Procedures, v4.0," 2022.
05. NSA and CISA, "5G Network Slicing: Security Considerations for Design, Deployment, and Maintenance," 2023.
06. Deloitte, "Connected Everything: Securing advanced connectivity use cases," 2023.
07. Deloitte, "Deloitte Cyber Threat Trends," 2023.
08. Deloitte, "5G Network Slicing: Security Considerations for Deign, Deployment and Maintenance," 2023.
09. Deloitte, "Emerging Technologies and Innovation: How 5G & IPv6 can enhance Edge AI solutions and shape the architecture of the future," 2023.

Deloitte.

Designed by CoReCreative Services. RITM1660565