AWS Whitepaper

Telco Lens

Telco Lens: AWS Whitepaper

Table of Contents

Abstract and introduction i
Telco-specific challenges 1
Industry context 1
Purpose of this lens 1
Target audience 2
Lens availability 2
Definitions 3
Design principles 4
Scenarios 6
5G core deployment on AWS 6
Reference architecture 7
Modernizing IMS networks on AWS 9
Reference architecture 10
Automatic lifecycle management of telco network functions 11
Reference architecture 12
Ingest NF packages into AWS TNB function catalog 12
Ingest network service descriptor (NSD) into AWS TNB network catalog 12
Create network instance (NI) and initiate deployment 13
Automated infrastructure provisioning 13
Automated NF installation and configuration 13
Ongoing lifecycle management 14
Operational excellence 15
Design principles 15
Organize 16
TELCOOPS01-BP01 Promote cross-functional collaboration across internal and external stakeholders 16
TELCOOPS01-BP02 Evaluate your governance and regulatory requirements 18
TELCOOPS01-BP03 Evaluate tradeoffs and threats while managing benefits and risks for telecommunication workloads 20
Multi-account strategy 22
TELCOOPS02-BP01 Telecommunication resources, operations and projects have identified owners 22
TELCOOPS02-BP02 Adopt a multi-account strategy to isolate different telecommunication workload 24

TELCOOPS02-BP03 Predict and mitigate deployment risks by adapting least disruptive deployment strategy for telecommunication workloads 27
Prepare 29
TELCOOPS03-BP01 Implement an enterprise-grade IP Address Management solution with cloud infrastructure integration 29
IP verification 31
TELCOOPS04-BP01 Implement the Bring Your Own IP (BYOIP) address processes and associate the prefixes with the IPAM solution 32
Operate 34
TELCOOPS05-BP01 Implement standardized provisioning and management of high-performance network interfaces 34
TELCOOPS05-BP02 Implement a structured sequence of interface assignment based on the persistent or ephemeral nature of the interfaces 37
Resource availability 39
TELCOOPS06-BP01 Verify resource availability for scale-out scenarios in telecommunication workloads 39
Geographical redundancy 41
TELCOOPS07-BP01 Design architecture patterns that maximize redundancy within available geographical constraints 42
Failure testing 44
TELCOOPS08-BP01 Develop and execute comprehensive failure testing programs that reflect cloud infrastructure characteristics 44
Performance baselining 46
TELCOOPS09-BP01 Implement comprehensive performance monitoring and baseline testing that accounts for shared infrastructure characteristics 46
Security 49
Design principles 49
Security foundations 50
TELCOSEC01-BP01 Follow the industry standards and local regulations to host telecom networks in the cloud, demonstrating a security commitment 51
TELCOSEC01-BP02 Encrypt data at-rest and sensitive traffic (control plane, data plane) using secure VPNs and custom encryption keys 53
Infrastructure protection 55
TELCOSEC02-BP01 Use secure control plane protocols between the network functions (NF) 56

TELCOSEC02-BP02 Implement features such as concealed or spoofed subscriber identities for signaling over the interfaces with access network 58
TELCOSEC02-BP03 Segment and isolate telco network domains and slices 60
External interface security 62
TELCOSEC03-BP01 Secure the APIs used to expose telco network functionality 62
TELSEC03-BP02 Deploy signaling firewall on the roaming and interconnecting interfaces with other telco networks 64
TELSEC03-BP03 Perform regular penetration testing on each of the protocols implemented on the signaling layer 66
Data Protection 68
TELCOSEC04-BP01 Enable encryption for CPNI and PII information at rest and in transit and access restriction 69
Reliability 72
Design principles 72
Foundations 73
TELCOREL01-BP01 Conduct regular load and failure tests to validate resilience on node, interface, and infrastructure levels 73
Workload architecture 76
TELCOREL02-BP01 Deploy user plane functions in a distributed architecture and highly available configuration 77
TELCOREL02-BP02 Implement full mesh between control plane and user plane functions 79
TELCOREL02-BP03 Implement a flexible network function (NF) design to leverage available infrastructure resources for autoscaling 81
TELCOREL02-BP04 Introduce an SCTP load balancer designed for control-plane network functions, carrier-grade performance, and high availability 83
TELCOREL02-BP05 Optimize failure recovery timers for the shared tenancy and potential for transient network issues in cloud environments 84
Change management 87
TELCOREL03-BP01 Use a controlled change management mechanism for software updates in production 87
TELCOREL03-BP02 Design a stateful, session-aware load balancing solution for cloud network functions (CNFs) to enable scaling and resilience 89
Failure management 92
TELCOREL04-BP01 Implement the Bypass Mode on peripheral network devices such as firewalls and network probes while connecting to the cloud 93

TELCOREL04-BP02 Implement dual network planes for signaling or control plane 94
TELCOREL04-BP03 Implement default configuration to bypass billing and charging services in case the system is down 96
Performance efficiency 99
Design principles 99
Network design principles 100
Architecture selection 100
TELCOPERF01-BP01 Use Edge locations to deploy latency sensitive telco network workloads such as RAN and user-plane nodes 101
TELCOPERF01-BP02 Select the infrastructure regions to deploy telco workloads based on performance requirements and regulatory considerations 103
TELCOPERF01-BP03 Deploy the control plane network functions on centralized locations to meet high scalability and agility requirements 106
Compute and hardware 108
TELCOPERF02-BP01 Implement containers to achieve optimal performance and resource utilization 108
Networking and content delivery 111
TELCOPERF03-BP01 Implement dedicated network infrastructure for control and user plane functions 111
TELCOPERF03-BP02 Deploy hardware acceleration solutions for enhanced packet processing and network performance 114
TELCOPERF03-BP03 Implement network timing synchronization and distribution mechanisms to verify service consistency 116
Process and culture 118
TELCOPERF04-BP01 Implement AI and ML-powered monitoring solutions to gain insights into network health and performance 119
TELCOPERF04-BP02 Monitor deviations from nominal performance parameters using system alerts and automated responses 121
Data management 124
TELCOPERF05-BP01 Implement data lifecycle management strategies for IVR and call logs to optimize storage costs and performance 124
Cost optimization 127
Design principles 127
Expenditure and usage awareness 127
TELCOCOST01-BP01 Implement attribution of cost to each telco domain hosted on the AWS (access, core, edge, OSS, and BSS) 128

Cost-effective resources 129
TELCOCOST02-BP01 Implement dynamic CNF sizing and scaling strategies based on actual subscriber demands and usage patterns 130
TELCOCOST02-BP02 Choose the most efficient compute resource for your Network Function 132
Data transfer 134
TELCOCOST03-BP01 Use edge-zones and services to implement cloud workloads made up of containers and microservices 134
Data storage 136
TELCOCOST04-BP01 Choose the appropriate type of storage for network functions backups, metrics, KPIs and the event records to reduce costs 137
TELCOCOST04-BP02 Use ETSI ENI based architectures to implement intelligent network slicing 139
Interoperability 141
TELCOCOST05-BP01 Explore open interface-based technology like RAN or vRAN to reduce network-related costs 141
Manage demand and supply resources 143
TELCOCOST06-BP01 Implement load-balancing techniques to achieve better utilization of hybrid network resources 143
Sustainability 146
Design principles 146
Energy optimization 148
TELCOSUS01-BP01 Implement energy-efficient infrastructure for telco networks 148
Data processing and storage 150
TELCOSUS02-BP01 Implement efficient data lifecycle management for telco networks 150
Investment protection 152
TELCOSUS03-BP01 Adopt circular economy principles for telco network assets 153
Climate change risk 154
TELCOSUS04-BP01 Enhance telco network resilience to climate change risks 155
Carbon tracking 157
TELCOSUS05-BP01 Establish comprehensive carbon footprint monitoring for telco networks 157
Conclusion 160
Contributors 162
Document revisions 163
AWS Glossary 164

Telco Lens - AWS Well-Architected

Publication date: **December 30, 2025** (*Document revisions*)

This paper describes the Telco Lens for the AWS Well-Architected Framework. The lens explores how to review and improve your cloud-based architecture for telecommunications workloads and better understand the impact of design decisions. We present general design principles and specific best practices aligned to the six pillars of the Well-Architected Framework.

Telco-specific challenges

The Telco Lens addresses unique operational challenges in telecommunications workloads:

- High traffic volumes with millions of transactions per second
- Low latency application and data retrieval responses with stringent service level agreements (SLAs)
- Rapid changes in traffic volumes and infrastructure scaling requirements
- Significant data transfer costs as a proportion of overall operational costs
- Need for cost efficiency and optimization due to low profit margins per transaction
- Strict end-to-end network latency requirements to meet industry service level objectives (SLOs)

Industry context

The telecommunications industry encompasses wireless and wireline service providers, independent software vendors (ISVs), and network equipment providers (NEPs) that deliver essential communication and connectivity services globally. This sector comprises companies transmitting data, voice, audio, and video worldwide.

Purpose of this lens

The Telco Lens provides:

1. Best practices for designing and operating telecommunications workloads in AWS
2. Guidance on building resiliency and sustainability into your AWS environment
3. Recommendations for performance, data privacy, and security

4. Strategies to assist communication service providers (CSPs) meet industry-specific requirements
5. Controls to expedite adoption of new services into your environment

Target audience

This document serves technology roles such as CTOs, architects, developers, and operations teams working on telco workloads. It will assist you to understand AWS best practices and strategies for designing and operating telco architectures aligned with the Well-Architected Framework.

Lens availability

Custom lenses extend the best practice guidance provided by AWS Well-Architected Tool. AWS WA Tool allows you to create your own custom lenses, or to use lenses created by others that have been shared with you.

To begin reviewing your telco workload, download and import the Telco Lens into AWS Well-Architected Tool from the public AWS Well-Architected custom lens GitHub repository.

Definitions

The following are common definitions used throughout the Telco Lens:

- **5G network:** The fifth generation of cellular network technology that provides significantly faster data speeds, lower latency, and higher capacity compared to previous 4G networks. 5G enables new use cases around the Internet of Things (IoT), augmented and virtual reality, autonomous vehicles, and more.
- **Network functions virtualization (NFV):** The concept of virtualizing network functions previously implemented in proprietary hardware, such as firewalls, load balancers, and routers. NFV allows these network functions to run on commercial off-the-shelf (COTS) hardware.
- **Software-defined networking (SDN):** The practice of decoupling the network control plane from the data forwarding plane, enabling dynamic, programmatic control of network behavior. SDN provides flexibility and agility in network management and service provisioning.
- **Edge computing:** The practice of processing data closer to the source of data generation, rather than in a centralized data center. Edge computing enables low-latency, high-bandwidth applications by reducing the distance data must travel.
- **Intelligent routing:** The use of advanced algorithms and real-time analytics to dynamically route network traffic based on factors like congestion, latency, security, and cost optimization.
- **Customer experience (CX) applications:** telco applications and solutions that enable an exceptional customer experience, such as self-service portals, chatbots, service assurance dashboards, and customer engagement tools.
- **Cloud network function (CNF):** A network function designed and implemented to run within containers, using cloud technologies and microservices architecture.
- **Network function (NF):** A functional building block within a network infrastructure, which has well-defined external interfaces and well-defined functional behavior.
- **Virtual network function (VNF):** Software implementation of network functions that can be deployed on virtual machines.
- **Control plane:** The part of a network that carries signaling traffic and is responsible for routing.
- **User plane or data plane:** The part of a network that carries user traffic.
- **Stream Control Transmission Protocol (SCTP):** A transport layer protocol that provides message-oriented communication with features for telecommunications signaling.
- **Communications Services Providers (CSP):** The network operators which build the networks according to the regulatory requirements and deliver telecommunication services to the users.

Design principles

The following design principles can assist you to achieve and maintain efficient telco workloads in the cloud:

- **Design for high availability and resilience:** Build fault-tolerant architectures with redundancy and failover mechanisms to provide continuous service availability. Use managed services and serverless offerings to reduce maintenance overhead and improve resilience. Implement automated infrastructure provisioning and recovery processes to quickly respond to and mitigate service disruptions.
- **Design for rapid scalability:** Engineer solutions that can dynamically scale compute, storage, and network resources to handle spikes in customer demand and traffic. Use auto scaling mechanisms and load balancing to efficiently distribute workloads across distributed infrastructure. Use cloud technologies like containers and serverless to enable rapid, on-demand scaling.
- **Design for operational efficiency:** Automate provisioning, deployment, monitoring, and remediation processes to improve operational agility. Implement infrastructure as code to enable consistent, repeatable, and auditable deployments. Use managed services to offload undifferentiated heavy lifting and focus on core business capabilities.
- **Design for data-driven insights:** Collect and analyze telemetry data from across the architecture to gain visibility into performance, usage, and customer behavior. Use advanced analytics and machine learning to uncover insights that inform ongoing architecture improvements. Empower developers and operators with self-service access to data to drive real-time decision-making.
- **Design for security:** Build security controls into the architecture from the ground up using the principle of least privilege. Implement robust identity and access management to control who and what can access sensitive resources. Continuously monitor for threats and vulnerabilities and rapidly respond to address them. Align the architecture with industry regulations and standards for data protection and privacy.
- **Design for cost optimization:** Continuously analyze usage patterns and cost drivers to identify opportunities for optimization. Use pricing models and purchasing options that align with workload requirements and usage patterns. Implement mechanisms to allocate and charge back costs to internal business units or customers.

The specific design principles and their emphasis may vary depending on the unique characteristics and priorities of the telco industry. It is important to tailor these principles to the specific context and requirements of the Telco Lens.

Scenarios

Communication service providers (CSPs) use public cloud services to address key business goals across four areas: networks transformation, customer experience management (including OSS and BSS), IT modernization, and innovative solutions for industry verticals.

This whitepaper focuses on the network's transformation area, highlighting three common telecom network implementation scenarios on AWS. The recommendations in the best practices are centered around these common scenarios.

Some CSPs use AWS Cloud services to build their on-demand disaster recovery networks, while others view cloud-based networks as an agile option to expand their existing network capacities.

In certain cases, greenfield CSPs choose AWS to deploy their first telco network, benefiting from reduced time to market and high scalability capabilities.

Although the network architecture is similar across these different cases, the specific architecture and services used vary depending on the part of the network being deployed. In this section, we present the assumptions made for each of these scenarios, the common drivers for the design, and a reference architecture for how these scenarios could be implemented.

Scenarios

- 5G core deployment on AWS
- Modernizing IMS networks on AWS
- Automatic lifecycle management of telco network functions
- Ingest NF packages into AWS TNB function catalog
- Ingest network service descriptor (NSD) into AWS TNB network catalog
- Create network instance (NI) and initiate deployment
- Automated infrastructure provisioning
- Automated NF installation and configuration
- Ongoing lifecycle management

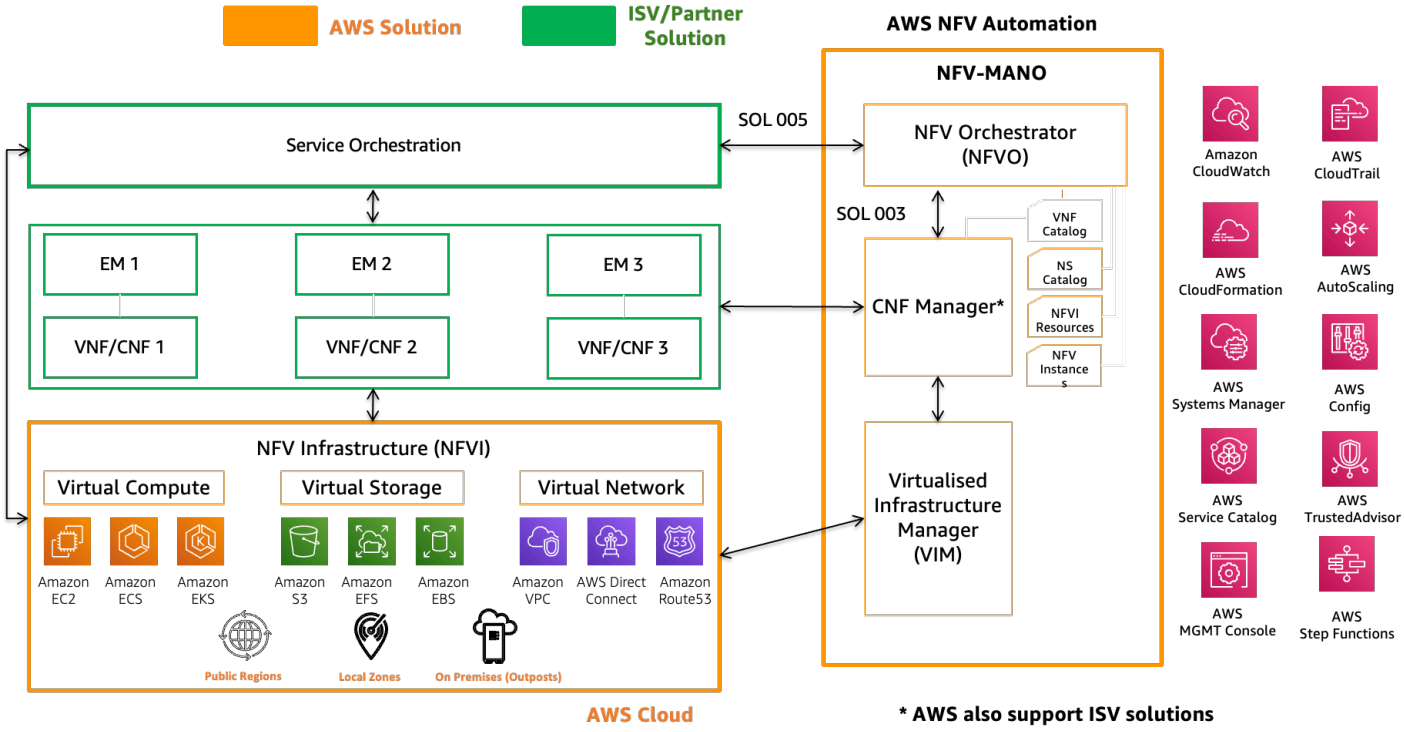
5G core deployment on AWS

This scenario explores deploying 5G core networks in the AWS Cloud. The 5G core is the center of 5G networks, providing key functionalities such as session management, user authentication,

and policy control. As telecommunications providers transition to 5G, migrating the 5G core to the cloud can offer significant benefits. This scenario covers topics such as virtualizing and containerizing 5G network functions and using AWS infrastructure services like AWS Outposts and AWS Local Zones to address performance and data residency requirements.

Key considerations include managing vendor dependencies, providing high availability and resiliency, and adopting the service-based architecture (SBA) for the 5G core. Practical advice includes evaluating current 5G core architectures, identifying opportunities to decouple and modernize components, and adopting DevOps practices for streamlined 5G core operations on AWS. By deploying the 5G core on AWS, telecommunications providers can unlock the agility, scalability, and cost optimization benefits of the cloud while future-proofing their networks for the evolving 5G landscape.

Reference architecture



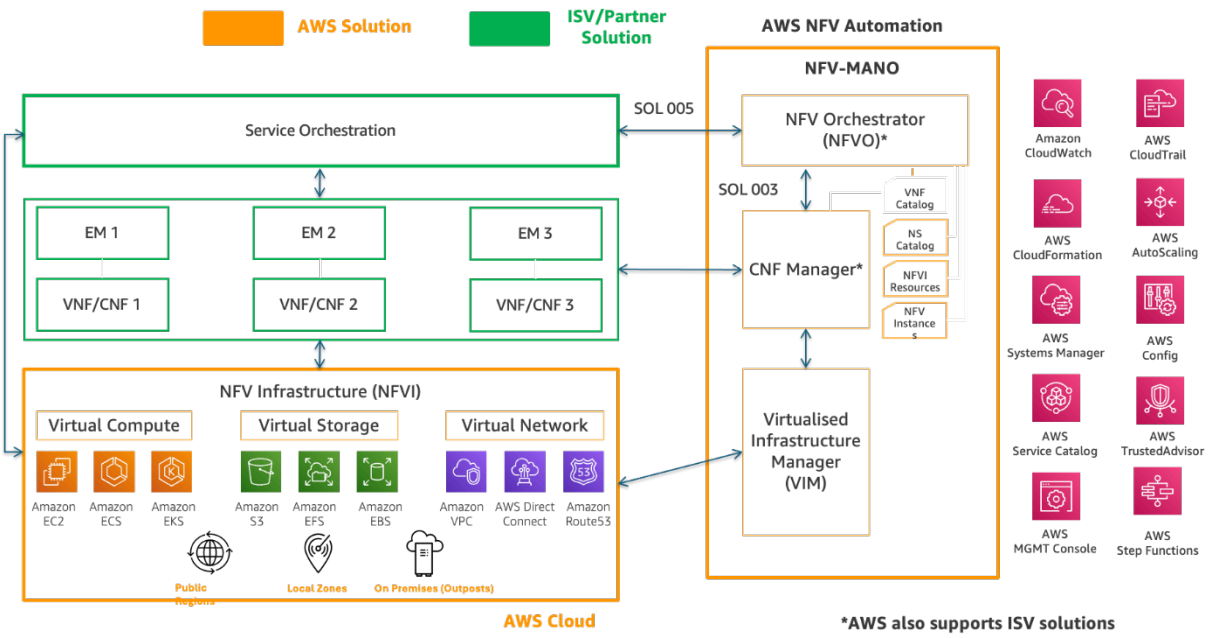


Figure 1. 5G core on globally available AWS infrastructure.

This reference architecture shows standard telecommunication architecture to deploy and manage virtual network functions (VNFs) or containerized network functions (CNFs). The set of AWS services can deliver the capabilities on the infrastructure and management layer. The part shown in green is provided by the ISV, which deploys these VNFs and CNFs using listed AWS services.

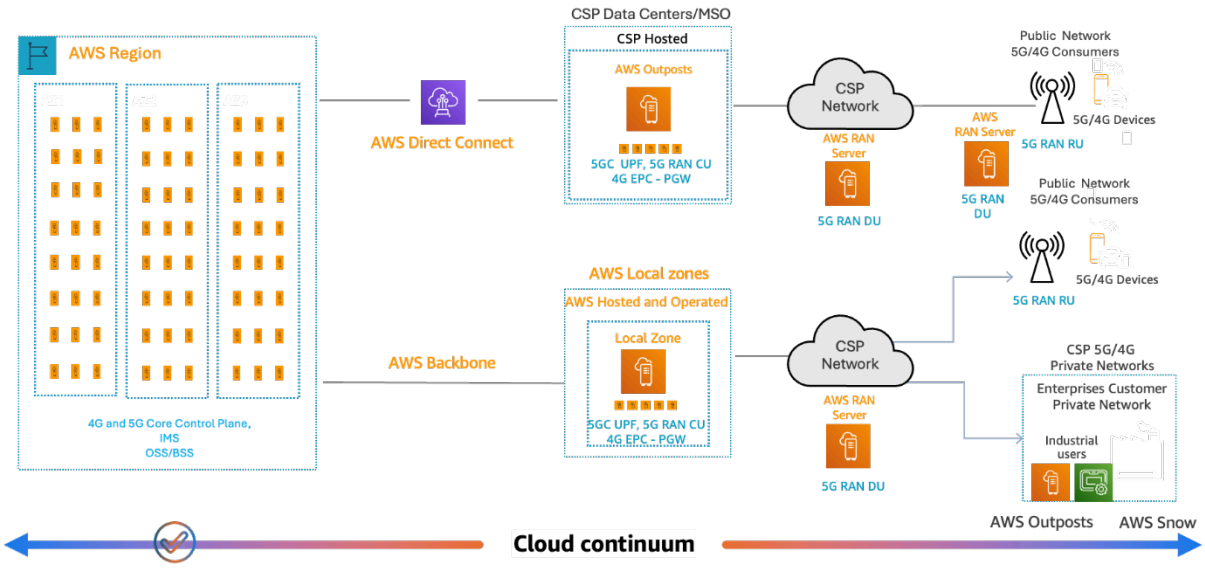


Figure 2. Bringing the cloud where the network needs it.

The network functions (NFs) in a telco core network can be split into control and user plane. The user plane NFs are throughput intensive and depend heavily on packet forwarding mechanisms.

However, the control plane functions are compute intensive and require higher levels of resiliency and reliability. In a distributed architecture, these control plane NFs are deployed in AWS Regions to offer better scalability and resilience. On the other hand, the user plane NFs are moved closer to the end user for better experience and low latency, using the higher throughput capabilities of Local Zones or AWS Outpost Racks. Outpost servers can be used to deploy the RAN part of the network as shown in the previous diagram.

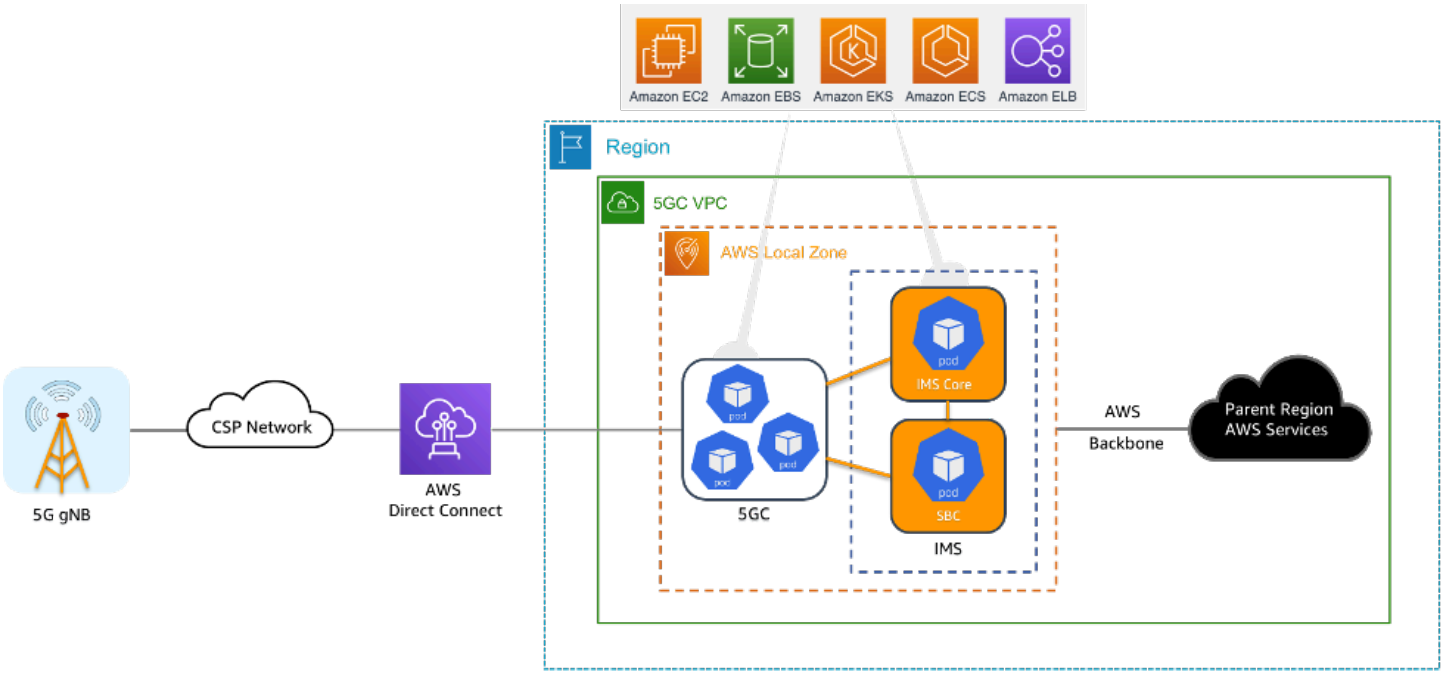
Modernizing IMS networks on AWS

This scenario explores modernizing IP multimedia subsystem (IMS) networks on the AWS Cloud. IMS is a key component of 5G networks, providing IP-based multimedia services like voice, video, and messaging. As telecommunications providers modernize their infrastructure for 5G, migrating legacy IMS deployments to the cloud can offer significant benefits.

This scenario covers topics such as virtualizing and containerizing IMS network functions, using AWS infrastructure services like Outposts and Local Zones to address latency and data residency requirements and implementing CI/CD pipelines using AWS developer tools for agile IMS deployments.

Key considerations include managing vendor dependencies, providing high availability and resiliency and transitioning towards a service-based architecture (SBA) for IMS. Practical advice includes evaluating current IMS architectures, identifying opportunities to decouple and modernize components, and adopting DevOps practices for streamlined IMS operations on AWS.

Reference architecture



AWS' global infrastructure of Regions and Availability Zones provides the foundation for deploying highly available and resilient IMS networks. Deploying IMS workloads across multiple AWS Regions offers benefits such as:

- **High availability and fault tolerance:** IMS components can be distributed across Availability Zones within a Region, or even across Regions, to provide continuous service in the event of a failure.
- **Disaster recovery and business continuity:** You can architect active-active or active-passive setups between Regions to enable fast recovery and failover in the case of a Regional disruption.
- **Scalability and elasticity:** The virtually unlimited capacity of AWS Regions assists to scale IMS resources up or down based on demand, improving overall system responsiveness.
- **Data sovereignty:** Strategically placing IMS data and workloads in specific AWS Regions can assist you to meet data residency and other regulatory requirements.
- Use AWS Global Accelerator and CloudFront to optimize network performance and reduce latency for geographically distributed IMS end users.
- Establish robust cross-Region monitoring, alerting, and automated remediation processes to improve the reliability and resilience of your global IMS deployment.

By incorporating the global AWS infrastructure and services, you can verify that your modernized IMS network meets the highest standards for availability, durability, and responsiveness, as outlined in the Reliability pillar of the AWS Well-Architected Framework.

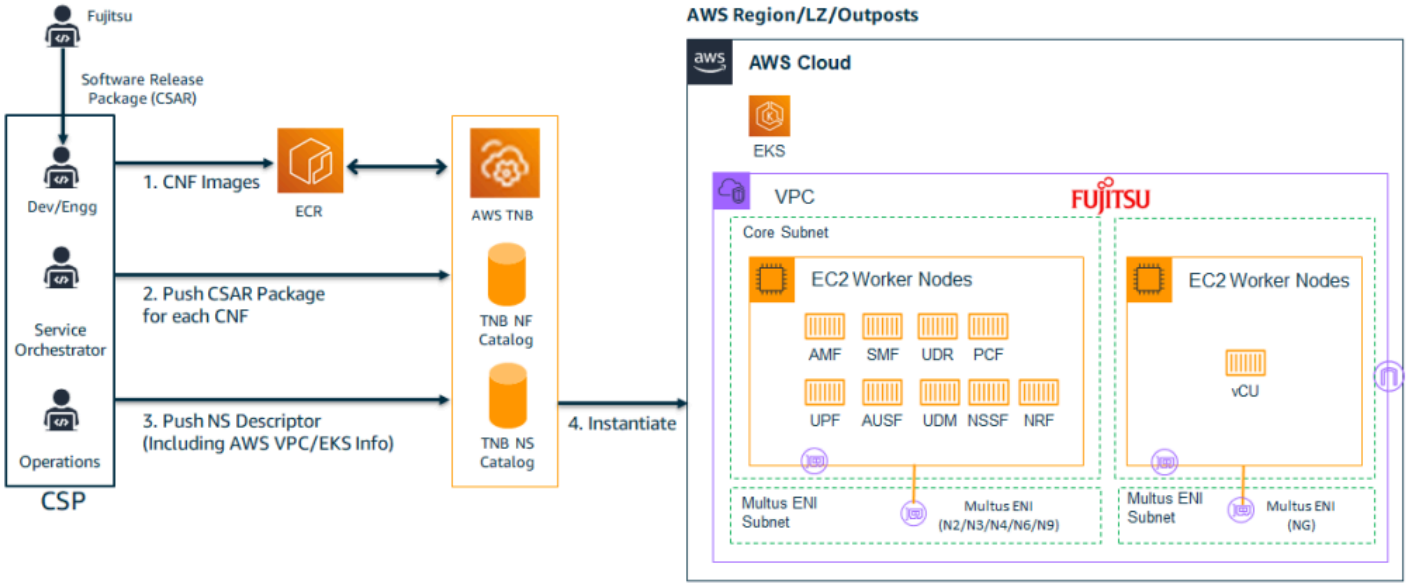
When modernizing IMS networks on AWS, verify the tight integration between the IMS, 5GC, and other network components to maintain end-to-end visibility and manageability. Jsing AWS monitoring and observability services like Amazon CloudWatch, AWS X-Ray, and AWS App Mesh can provide comprehensive insights into IMS workload performance.

Implement robust security controls such as AWS IAM, VPCs, security groups, and network ACLs to protect critical IMS assets. Plan disaster recovery and business continuity, using features like multi-AZ database deployments and AWS' global infrastructure. Continuous optimization of resource utilization, scaling, and load balancing based on traffic patterns and metrics is crucial. Establish a proactive maintenance and patching cadence for OS, middleware, and application components to maintain a secure and reliable IMS environment. Finally, implement automated testing and canary deployments to mitigate the risk of changes to these critical telecommunications systems.

Automatic lifecycle management of telco network functions

This scenario explores the deployment of end-to-end 5G networks, including radio access network (RAN) and 5G core, using the AWS telco Network Builder (TNB) service. It addresses the challenges faced by CSPs in unifying their operating model and automating network deployments across the 5G environment, including RAN, core, and edge infrastructure.

Reference architecture



Fujitsu 5G network deployment using AWS TNB

Ingest NF packages into AWS TNB function catalog

CSPs ingest their 5G network function (NF) packages (for example, vCU, vUPF and vAMF) into the AWS TNB function catalog. These packages are in the form of a Cloud Service Archive (CSAR) file containing the NF descriptor, Helm charts, and custom scripts.

Recommendation: Work closely with NF vendors to verify the packages adhere to cloud design principles and meet the packaging requirements of AWS TNB. Establish a CI/CD pipeline to automate the ingestion of updated NF packages.

Practical advice: Thoroughly test the NF packages in a non-production environment before ingesting them into the production catalog. Maintain version control and cataloging of the NF packages to enable rollbacks and updates.

Ingest network service descriptor (NSD) into AWS TNB network catalog

CSPs ingest the NSD, which describes the required compute and network resources, as well as the NFs to be deployed, into the AWS TNB network catalog.

Recommendation: Design the NSD to be modular and reusable, allowing for the creation of multiple network instances from a single template. Leverage AWS TNB's support for ETSI SOL003/SOL005 APIs to integrate with existing ETSI-based service orchestrators.

Practical advice: Thoroughly test the NSD in a non-production environment to verify the correct mapping of resources and NFs. Maintain version control of the NSD to enable updates and rollbacks.

Create network instance (NI) and initiate deployment

CSPs create a new network instance (NI) using the previously ingested NSD and then initiate the deployment process.

Recommendation: Implement a standardized approach for creating NIs to verify consistency and repeatability across deployments. Leverage AWS TNB's ability to create multiple NIs from a single NSD template to support use cases like private network deployments.

Practical advice: Monitor the deployment progress and status using the AWS TNB console or APIs. Implement automated triggering of the deployment process as part of the CI/CD pipeline.

Automated infrastructure provisioning

AWS TNB automatically provisions the required compute and network resources, such as VPCs, subnets, Elastic Network Interfaces (ENIs), Amazon EKS clusters, and Amazon EC2 instances, to support the 5G NF deployment.

Recommendations: Verify the network and compute resources are designed to meet the performance, scalability, and resiliency requirements of the 5G network. Use AWS services like VPC, Amazon EKS, and Amazon EC2 to benefit from the built-in security, monitoring, and management capabilities.

Practical advice: Implement robust access controls, network segmentation, and security monitoring to protect the 5G network infrastructure. Establish automated remediation processes to address infrastructure-related issues.

Automated NF installation and configuration

AWS TNB performs the installation and configuration of the 5G NFs (for example, CU, AMF, UPF, and SMF) on the provisioned infrastructure.

Recommendations: Verify the NF packages and Helm charts are properly designed to use the capabilities of Kubernetes and the underlying cloud infrastructure. Implement automated testing and canary deployments to validate the NF installations.

Practical advice: Monitor the NF deployments for errors or issues and leverage the AWS TNB APIs to programmatically retrieve logs and metrics for troubleshooting. Establish automated rollback and recovery procedures in case of deployment failures.

Ongoing lifecycle management

AWS TNB provides a unified view and control over the deployed 5G network, enabling CSPs to perform lifecycle management operations such as scaling, updating, and terminating network instances.

Recommendations: Implement comprehensive monitoring, observability, and automated remediation processes to verify the reliability and resilience of the 5G network. Leverage AWS services like CloudWatch, X-Ray, and App Mesh to gain insights into network performance and health.

Practical advice: Establish a proactive maintenance and patching cadence for the 5G network components to maintain security and operational integrity. Regularly review resource utilization and scaling metrics to optimize the network for cost and performance.

By following these reference architecture and recommendations, CSPs can leverage the AWS telco Network Builder to automate the deployment and lifecycle management of their 5G RAN and Core networks, unlocking the benefits of agility, scalability, and cost optimization while maintaining the reliability and resilience of their 5G infrastructure.

Operational excellence

This section describes the operational excellence pillar for the AWS Well-Architected Telco Lens. It provides AWS customers with a set of best practices and guidance on how telecommunications service providers can achieve operational excellence in their AWS cloud environments.

Operational excellence is a commitment to build software correctly while consistently delivering a great customer experience. The operational excellence pillar contains best practices for organizing your team, designing your workload, operating it at scale, and evolving it over time.

Focus areas

- Design principles
- Organize
- Multi-account strategy
- Prepare
- IP verification
- Operate
- Resource availability
- Geographical redundancy
- Failure testing
- Performance baselining

Design principles

The AWS Well-Architected Framework identifies a set of general design principles to facilitate superior design in the cloud. In addition, the following design principles should also be considered for designing and operating telco workloads:

- **Cross-functional collaboration**: Foster collaboration between network engineers, cloud architects, ISV architects, and security experts to verify efficient communication and knowledge sharing across the organization.
- **Governance integration**: Systematically assess and verify telco governance frameworks, regulatory mandates, and security requirements across network operations.

- **Risk-based decision making**: Conduct comprehensive analysis of benefits, risks, and tradeoffs associated with telco workloads to make informed architectural and operational decisions.
- **Multi-account isolation with interoperability**: Implement a strategic multi-account architecture that effectively separates and isolates telecommunications workloads while maintaining necessary interconnectivity and operational efficiency.
- **Observability-driven operations**: Design and implement comprehensive observability, traceability, and monitoring frameworks aligned with telecommunications workload requirements and business objectives.
- **Manage IP addressing effectively**: Implement cloud-native IP address management (IPAM) solutions optimized for telco workloads requiring secondary network interfaces.
- **Verify resource availability**: Plan for resource availability for scale-out scenarios in telecommunication workloads, considering both new service and run-rate capacity planning.

Organize

TELCOOPS01: How do you create the optimal cross-functional team?

Telecom products frequently interface with multiple upstream and downstream applications, interacting with both internal and external stakeholders. Creating cross-functional teams is critical to gain comprehensive visibility across end-to-end workflows.

Best practices

- TELCOOPS01-BP01 Promote cross-functional collaboration across internal and external stakeholders
- TELCOOPS01-BP02 Evaluate your governance and regulatory requirements
- TELCOOPS01-BP03 Evaluate tradeoffs and threats while managing benefits and risks for telecommunication workloads

TELCOOPS01-BP01 Promote cross-functional collaboration across internal and external stakeholders

Involve key stakeholders, including business, development, and operations teams, to determine where to focus efforts on external and internal goals. Identify long-term and short-term goals

which assist to prioritize tasks with added efficiency. Encourage collaboration between different teams, such as network engineers, cloud architects, ISV architects and security experts, to verify efficient communication and knowledge sharing across the organization.

Desired outcome:

- Established cross-functional teams with clear roles and responsibilities.
- Improved communication and collaboration between network, cloud, security, and business teams.
- Aligned short-term and long-term goals across different stakeholder groups.
- Enhanced decision-making through diverse expertise and perspectives.
- Faster problem resolution and innovation cycles.

Common anti-patterns:

- Siloed teams working in isolation without regular interaction.
- Lack of shared goals and metrics across teams.
- Communication barriers between technical and business stakeholders.
- Insufficient involvement of security and compliance-aligned teams in preliminary stages.
- No clear escalation paths for cross-team issues.
- Collaboration without structured processes.
- Missing documentation of decisions and rationales.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Creating effective cross-functional teams requires a structured approach that balances organizational needs with individual expertise. Begin by establishing clear team charters that define the scope, objectives, and decision-making authority for each cross-functional group. Implement regular synchronization mechanisms such as daily stand-ups, weekly coordination meetings, and monthly strategic reviews to verify continuous alignment across teams. Create standardized communication channels and documentation practices that enable transparent information sharing while maintaining clear escalation paths for critical issues. Consider using agile methodologies adapted for telecommunications environments to improve collaboration efficiency and response times to changing requirements.

Implementation steps

- Use AWS Organizations to create organizational units (OUs) that reflect team structures, and AWS IAM Identity Center for centralized access management and role-based permissions.
- Implement AWS Systems Manager OpsCenter for centralized operations management and integrate collaboration tools through Amazon EventBridge for automated notifications.
- Use Service Catalog to standardize approved services and configurations across teams, with AWS tags for resource tracking and ownership.
- Deploy AWS Systems Manager Change Manager for standardized change processes and Amazon CloudWatch for automated operational dashboards.
- Use AWS Systems Manager Automation for process standardization and AWS Config for resource monitoring.

Resources

Key AWS services:

- AWS Organizations
- AWS Identity and Access Management (IAM)
- AWS Resource Access Manager (RAM)
- AWS Systems Manager
- Service Catalog

TELCOOPS01-BP02 Evaluate your governance and regulatory requirements

Systematically assess and verify telco governance frameworks, regulatory mandates, and security requirements across network operations. This includes evaluating legal intercept capabilities, security controls, and industry-specific regulations while maintaining documentation of assessment activities and findings. Regular evaluation verifies continued alignment with evolving regulatory landscapes and identify gaps that require remediation.

Desired outcome:

- Comprehensive understanding of regulatory requirements.

- Clear governance frameworks aligned with industry standards.
- Documented controls and procedures.
- Regular assessment and verification processes.
- Traceability of activities.
- Proactive identification of regulatory changes.

Common anti-patterns:

- Reactive approach to requirements.
- Incomplete documentation of regulatory obligations.
- Missing or outdated controls.
- No regular assessments.
- Lack of monitoring tools.
- Insufficient training on regulatory requirements.
- Random governance processes.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Establish a comprehensive governance framework that addresses both industry-specific regulations and organizational policies. Create a systematic approach for identifying, documenting, and tracking regulatory requirements using a centralized management system that maintains status and upcoming changes. Implement regular assessment cycles that include both internal audits and external validations to verify continued adherance with telecommunications regulations and standards. Develop a robust documentation system that maintains evidence of activities, including regular testing of controls, training records, and audit trails for regulatory reporting.

Implementation steps

- Deploy AWS Audit Manager to evaluate adherence with regulatory standards and AWS Config to assess resource configurations against rules.
- Implement AWS Control Tower for multi-account governance and AWS Organizations for policy management through Service Control Policies (SCPs).

- Configure AWS Security Hub CSPM for centralized security monitoring and AWS CloudTrail for comprehensive API activity logging.
- Use AWS Systems Manager Documents to create standardized procedures and AWS IAM Access Analyzer for continuous permission validation.
- Deploy Amazon EventBridge for automated checks and AWS Config Rules for ongoing configuration assessment.

Resources

Key AWS services:

- AWS Config
- AWS Audit Manager
- AWS Security Hub CSPM
- AWS Control Tower
- AWS CloudTrail

TELCOOPS01-BP03 Evaluate tradeoffs and threats while managing benefits and risks for telecommunication workloads

Conduct comprehensive analysis of benefits, risks, and tradeoffs associated with telco workloads to make informed architectural and operational decisions. This involves evaluating multiple dimensions including technical capabilities, security implications, vendor relationships, and organizational impact. The assessment should balance immediate operational needs with long-term strategic objectives while considering security threats, technological maturity, and cross-functional requirements.

Desired outcome:

- Balanced risk-benefit analysis framework.
- Clear understanding of technical and operational tradeoffs.
- Documented threat assessment methodology.
- Risk mitigation strategies.
- Quantified impact analysis.

- Strategic alignment with business objectives.

Common anti-patterns:

- Making decisions without proper risk analysis.
- Ignoring technical debt implications.
- Lack of threat modeling.
- Insufficient stakeholder input in risk decisions.
- No documented decision criteria.
- Overlooking long-term consequences.
- Missing risk registers.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Develop a structured risk assessment framework that considers both technical and business impacts of architectural decisions in telecommunications environments. Implement a systematic approach to threat modeling that includes regular security assessments, vulnerability scanning, and penetration testing to identify potential risks early in the development cycle. Create a decision-making framework that weighs multiple factors including performance requirements, security implications, operational overhead, and cost considerations to verify balanced outcomes. Establish regular review cycles to reassess previous decisions and adjust strategies based on changing threat landscapes and business requirements.

Implementation steps

- Create a standardized methodology for evaluating risks, threats, and opportunities across telecommunication workloads.
- Conduct comprehensive threat modeling and vulnerability assessments with documented mitigation strategies.
- Establish quantifiable metrics for evaluating potential benefits and ROI of architectural decisions and operational changes.
- Implement structured decision frameworks with clear evaluation criteria and approval workflows.
- Establish regular review cycles to assess the effectiveness of decisions and adjust strategies based on outcomes.

Resources

Related services:

- AWS Security Hub CSPM
- Amazon GuardDuty
- AWS WAF
- Amazon Inspector
- AWS Trusted Advisor
- Amazon Detective

Multi-account strategy

TELCOOPS02: What factors should be considered when implementing a multi-account strategy for telecommunication workloads?

This addresses the critical factors and best practices for designing and implementing a multi-account architecture that effectively supports telecommunications operations while maintaining security and operational efficiency.

Best practices

- TELCOOPS02-BP01 Telecommunication resources, operations and projects have identified owners
- TELCOOPS02-BP02 Adopt a multi-account strategy to isolate different telecommunication workload
- TELCOOPS02-BP03 Predict and mitigate deployment risks by adapting least disruptive deployment strategy for telecommunication workloads

TELCOOPS02-BP01 Telecommunication resources, operations and projects have identified owners

Establish clear ownership and accountability frameworks for telecommunications assets, operations, and projects through formal assignment of responsibilities to specific individuals

or teams. This includes mapping ownership of applications, VNFs, solutions, and infrastructure components while documenting their business value proposition and justification for ownership allocation. Clear ownership structures enable efficient decision-making, streamlined problem resolution, and effective resource management across the telecommunications environment.

Desired outcome:

- Clear ownership structure for resources.
- Documented responsibilities and accountabilities.
- Efficient resource allocation.
- Streamlined decision-making process.
- Effective resource management.
- Clear escalation paths.

Common anti-patterns:

- Undefined resource ownership.
- Shared responsibilities without clear boundaries.
- Missing accountability frameworks.
- Incomplete resource documentation.
- Unclear decision authority.
- Resource orphaning.
- Ambiguous escalation paths.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Establish a comprehensive resource ownership model that clearly defines responsibilities for telecommunications assets, operations, and projects. Implement a centralized resource management system that tracks ownership details, including primary and secondary owners, escalation paths, and relevant stakeholders. Create detailed RACI (responsible, accountable, consulted, informed) matrices for each major resource category to verify clear understanding of roles and decision-making authority. Develop standardized handover procedures and documentation requirements to maintain continuity during ownership transitions or organizational changes.

Implementation steps

- Implement AWS Resource Groups and Tag Editor to define and manage resource ownership tags, and integrate with AWS Organizations for hierarchical resource management.
- Use AWS Config to maintain resource inventory and relationships and AWS Systems Manager Resource Groups for logical grouping of resources by ownership.
- Deploy Service Catalog for standardized resource provisioning with predefined ownership tags and AWS CloudFormation for automated resource creation with ownership metadata.
- Configure AWS Config Rules for ownership monitoring and Amazon CloudWatch for resource utilization tracking by owner.
- Use AWS Systems Manager OpsCenter for ownership-related operations management and AWS Resource Access Manager for controlled resource sharing.

Resources

Key AWS services:

- Resource Groups and Tagging for AWS
- AWS Organizations
- Service Catalog
- AWS Systems Manager

TELCOOPS02-BP02 Adopt a multi-account strategy to isolate different telecommunication workload

Implement a strategic multi-account architecture that effectively separates and isolates telecommunications workloads while maintaining necessary interconnectivity and operational efficiency. This approach creates logical boundaries between different environments, services, and data classifications to enhance security and operational stability. The strategy should balance the benefits of isolation with the need for seamless integration and management across the telecommunications landscape.

Desired outcome:

- Secure workload isolation.
- Optimized resource management.

- Clear account boundaries.
- Efficient cross-account connectivity.
- Standardized account governance.
- Scalable account structure.

Common anti-patterns:

- Single account for each workload.
- No account categorization.
- Excessive account fragmentation.
- Missing connectivity strategy.
- Inconsistent security controls.
- Ad-hoc account creation.
- Poor resource sharing controls.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Design a multi-account architecture that addresses telecommunications-specific isolation requirements.

Categorize workloads based on:

- Regulatory data sovereignty (separate accounts per country where laws mandate subscriber data remain within borders like EU GDPR, cybersecurity laws, and data localization regulations).
- Subscriber data sensitivity (isolated accounts for CPNI and PII processing with enhanced security controls versus anonymized network telemetry).
- Network function security domains (dedicated accounts for control plane functions like AMF or SMF requiring high security, user plane UPF functions optimized for throughput, and signaling functions requiring Diameter or SIP/SS7 firewall protection).

This categorization enables mapping telecommunications regulatory and operational requirements to appropriate account boundaries while maintaining cloud infrastructure efficiency.

Implement a hierarchical Organizational Unit (OU) structure reflecting your telecommunications architecture:

- RAN OU for edge-deployed radio functions
- 5G core OU with separated control and user plane accounts
- IMS OU for multimedia subsystem functions
- BSS OU for billing (PCI DSS compliance) and CRM (PII protection)
- OSS OU for network management and assurance
- Regional compliance OUs for per-country data sovereignty

Apply service control policies enforcing regulatory requirements (deny cross-region replication from EU accounts for GDPR, mandatory encryption for billing accounts). Establish standardized connectivity using centralized transit and private connectivity services with separate paths for control plane signaling versus user plane data traffic. Deploy centralized logging, monitoring, and security tooling in dedicated Security OU accounts to maintain visibility and unified governance across telco workloads.

Implementation steps

- Deploy AWS Control Tower for automated account setup and use AWS Organizations for hierarchical account structure and policy management.
- Implement AWS IAM Identity Center for centralized access management and AWS Security Hub CSPM for multi-account security monitoring.
- Configure AWS Transit Gateway for centralized network connectivity and AWS Network Firewall for consistent security controls.
- Use AWS Resource Access Manager for cross-account resource sharing and AWS Systems Manager for centralized operations management.
- Deploy AWS CloudFormation StackSets for multi-account resource deployment and AWS Organizations for policy-based governance.

Resources

Key AWS services:

- AWS Organizations

- AWS Control Tower
- AWS Transit Gateway
- AWS Resource Access Manager
- AWS IAM Identity Center
- AWS Network Firewall

TELCOOPS02-BP03 Predict and mitigate deployment risks by adapting least disruptive deployment strategy for telecommunication workloads

Implement risk-aware deployment strategies that minimize service disruption in telecommunications environments through predictive analysis, parallel testing, and proactive capacity management. This approach combines advanced analytics, infrastructure redundancy, and staged deployment methodologies to verify seamless service delivery while implementing necessary changes and updates to the network infrastructure.

Desired outcome:

- Minimized service disruption during deployments.
- Predictable deployment outcomes.
- Risk-aware deployment processes.
- Automated rollback capabilities.
- Efficient change management.
- Service continuity maintenance.

Common anti-patterns:

- Big bang deployments.
- No rollback planning.
- Insufficient testing.
- Missing impact analysis.
- Poor deployment timing.
- Inadequate capacity planning.
- No deployment validation.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Design and implement deployment strategies that minimize service disruption while maintaining successful changes in telecommunication environments. Establish a comprehensive risk assessment framework that evaluates potential impacts before deployments, including dependencies, customer impact, and resource requirements. Create deployment patterns that utilize blue-green deployments, canary releases, or rolling updates to minimize service interruption while maintaining the ability to quickly rollback changes. Implement automated validation and testing procedures throughout the deployment pipeline to catch potential issues early and verify service quality is maintained during and after deployments.

Implementation steps

- Use AWS Systems Manager Change Manager for change risk evaluation and AWS Config for pre-deployment checks.
- Implement AWS CodeDeploy for blue-green deployments and AWS CloudFormation for infrastructure deployment with rollback capabilities.
- Deploy AWS CodePipeline for automated testing workflows and AWS Fault Injection Service for controlled failure testing.
- Configure Amazon CloudWatch for deployment monitoring and AWS X-Ray for deployment impact tracing.
- Use AWS Systems Manager Automation for standardized deployment procedures and Service Catalog for approved deployment patterns.

Resources

Key AWS services:

- AWS CodeDeploy
- AWS CloudFormation
- AWS Systems Manager
- Amazon CloudWatch
- AWS App Runner
- AWS Fault Injection Service

Prepare

TELCOOPS03: How do you effectively manage IP address allocation and assignment for telco and high-performance networking workloads?

Many telco and high-performance networking workloads, such as those using user-plane separation, DPDK, or SR-IOV, require secondary network interfaces at both the infrastructure and container levels. These workloads need robust IP address management solutions that integrate seamlessly with cloud infrastructure while providing enterprise-grade support and maintenance. The solution must address operational requirements including high availability, performance, scalability, and support for cloud-native architectures.

Best practices

- TELCOOPS03-BP01 Implement an enterprise-grade IP Address Management solution with cloud infrastructure integration

TELCOOPS03-BP01 Implement an enterprise-grade IP Address Management solution with cloud infrastructure integration

Implement a cloud-based IPAM solution that is optimized for performance and tightly integrated with the underlying cloud networking infrastructure. This should provide reliable, scalable, and high-throughput IP address management capabilities tailored for telco and high-performance networking workloads that require secondary network interfaces.

Desired outcome:

- Reliable and scalable IP address management for secondary interfaces.
- Seamless integration with cloud infrastructure and container solutions.
- Enterprise-level support and maintenance.
- Automated IP allocation and de-allocation.
- Clear upgrade and patch management path.
- Consistent performance at scale.

Common anti-patterns:

- Using unsupported or community-maintained IPAM solutions.
- Implementing manual IP address management processes.
- Missing integration with cloud infrastructure.
- Lack of monitoring and alerting for IP management.
- No consideration for support and maintenance.
- Poor documentation and operational procedures.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement an IP address management solution that meets enterprise requirements while maintaining cloud-based integration capabilities. Focus on solutions that are officially supported by both the solution vendor and container distribution provider. Verify the selected solution provides comprehensive monitoring, automation capabilities, and clear escalation paths for support. Consider future integration possibilities with native cloud services while maintaining operational excellence today.

Implementation steps

- Solution evaluation:
  - Document requirements for IP management including scale, performance, and support needs.
  - Evaluate enterprise-supported solutions from solution providers.
  - Assess integration capabilities with AWS services.
  - Verify support and maintenance agreements.
  - Review upgrade and patch processes.
- Integration design:
  - Design integration architecture with AWS networking services.
  - Plan monitoring and observability integration.
  - Create automation workflows for IP management.
  - Document operational procedures and support processes.
- Implementation and testing:
  - Deploy solution in test environment.
  - Validate integration with AWS services.

    - Perform scale and performance testing.
    - Verify monitoring and alerting capabilities.
    - Test upgrade and patch procedures.
- Operational readiness:
    - Create operational runbooks .
    - Establish support procedures and escalation paths.
    - Document backup and recovery processes.
    - Implement change management procedures.
- Continuous improvement:
    - Regular review of solution performance.
    - Monitor for new native cloud capabilities.
    - Maintain upgrade and patch adherence.
    - Regular testing of operational procedures.

Resources

Key AWS services:

- Amazon VPC
- AWS Transit Gateway
- Amazon CloudWatch
- AWS Systems Manager
- AWS Config
- AWS CloudTrail

IP verification

TELCOOPS04: How do you represent IP addresses to the internet and verify the quality of the address representation?

Telco's own large CIDR blocks of IP addresses and wish to have them represented with their ownership even when advertised by AWS. The IP address CIDR block are assigned to telco resources in the cloud and should be propagated through AWS to the Internet where both the CIDR block and the ASN associated with the prefixes are represented as owned by the telco. This improves the reputation of the address prefix propagated across the Internet for telco subscribers and gives CSPs responsibility and accountability of the address space.

Best practices

- TELCOOPS04-BP01 Implement the Bring Your Own IP (BYOIP) address processes and associate the prefixes with the IPAM solution

TELCOOPS04-BP01 Implement the Bring Your Own IP (BYOIP) address processes and associate the prefixes with the IPAM solution

The IPAM is a shared resource that can be provisioned in a single Region in a shared services account. Pools of IP addresses can be shared using Resource Access Manager to sub-accounts and specific Regions where telco assets reside. The BYOIP process allows the telco to present their mobile subscribers to the Internet as members of their own network instead of AWS. The size of the address pools should be forecasted with sufficient room for growth. It is important to minimize the fragmentation of the address blocks and allow for contiguous blocks for a given Region. This will simplify access controls such as firewall rules, NACLs, and security groups. Finally, the bring your own ASN process should be invoked to enable AWS to use BGP to represent the BYOIP pool to the Internet as owned by the CSPs.

Desired outcome:

- Controlled IP address representation.
- Maintained IP reputation.
- Efficient address pool management.
- Proper ASN association.
- Optimized address utilization.
- Clear ownership documentation.

Common anti-patterns:

- Uncontrolled address advertising.

- Poor address pool planning.
- Missing ASN documentation.
- Fragmented address spaces.
- Inefficient address allocation.
- No address forecasting.
- Inconsistent routing policies.

Level of risk if the best practice is not established: High

Implementation guidance

Design a comprehensive BYOIP implementation strategy that maintains control over IP address representation while maintaining proper routing and advertisement. Establish a detailed IP address management plan that includes address pool allocation, subnet design, and forecasting mechanisms to support future growth without fragmentation. Implement robust BGP routing policies and ASN configurations that maintain proper route advertisement while maintaining optimal path selection and failover capabilities. Create detailed documentation and operational procedures for managing IP address assignments, including regular audits of address utilization and routing effectiveness to maintain the integrity of your network addressing scheme.

Implementation steps

- Use Amazon VPC IPAM for address pool management and AWS Resource Groups for IP address organization.
- Configure AWS Global Accelerator for IP address advertising and Amazon Route 53 for DNS management.
- Set up AWS Direct Connect for BGP routing and AWS Transit Gateway for route propagation.
- Deploy Amazon VPC IPAM for BYOIP address management and AWS Network Manager for network visibility.
- Implement AWS Systems Manager for operational tasks and Amazon CloudWatch for IP address monitoring.

Resources

Key AWS services:

- Use VPC IP Address Manager to manage subnet CIDRs
- AWS Direct Connect
- Amazon Route 53
- AWS Transit Gateway
- AWS Network Manager
- AWS Global Accelerator

Operate

TELCOOPS05: How do you automate the provisioning and assignment of a secondary network interface?

This generic question captures the key challenge without referencing a specific cloud provider or networking technology like ENA. The associated best practice and description can also be updated to be more cloud-agnostic.

Best practices

- TELCOOPS05-BP01 Implement standardized provisioning and management of high-performance network interfaces
- TELCOOPS05-BP02 Implement a structured sequence of interface assignment based on the persistent or ephemeral nature of the interfaces

TELCOOPS05-BP01 Implement standardized provisioning and management of high-performance network interfaces

telco and high-performance network workloads require specialized networking interfaces (like SR-IOV, DPDK, and Multus) for separating control, management, and data plane traffic. While other cloud and container solutions provide native support for these interfaces, AWS currently requires custom implementation. This creates challenges for standardization and automation in large-scale deployments. Until AWS support becomes available, organizations need to implement enterprise-grade solutions that provide consistent, supported, and maintainable approaches to interface provisioning and management.

Desired outcome:

- Automated interface provisioning.
- Consistent configuration.
- Scalable deployment process.
- Efficient resource utilization.
- Standardized networking setup.
- Reliable interface management.

Common anti-patterns:

- Manual interface configuration.
- Inconsistent setup procedures.
- Poor documentation.
- No automation framework.
- Missing standardization.
- Inefficient resource allocation.
- Ad-hoc management.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Develop an automated interface provisioning system that maintains consistent configuration and optimal performance for telecommunication workloads. Create standardized templates and workflows that handle the entire lifecycle of network interfaces, from initial provisioning through configuration, monitoring, and decommissioning. Implement comprehensive validation checks at each stage of the provisioning process to verify proper configuration and block common issues such as IP conflicts or misconfigurations. Establish monitoring and alerting mechanisms that track interface health, performance metrics, and utilization patterns to enable proactive management and optimization.

Implementation steps

1. Solution selection:

- Evaluate enterprise-supported container networking solutions
- Verify integration capabilities with AWS infrastructure
- Verify vendor support and maintenance agreements
- Document migration strategy for future native AWS capabilities

2. Interface management:
   - Implement vendor-supported automation tools
   - Use solution-provided operators where available
   - Integrate with AWS service limits and quotas
   - Maintain consistent interface naming and tagging
3. Operational integration:
   - Deploy vendor-recommended monitoring solutions
   - Integrate with Amazon CloudWatch for infrastructure monitoring
   - Establish clear operational boundaries between systems and AWS
   - Implement automated health checks
4. Support and maintenance:
   - Establish clear support channels with both vendor and AWS
   - Document operational procedures
   - Maintain upgrade and patch procedures
   - Regular testing of failover scenarios
5. Future readiness:
   - Monitor AWS feature announcements
   - Maintain modularity in automation
   - Document transition strategy
   - Regular architecture reviews

Resources

Key AWS services:

- AWS CloudFormation
- AWS Lambda
- Amazon EventBridge

- AWS Systems Manager
- AWS Cloud Development Kit (AWS CDK) (CDK)
- AWS Network Manager

TELCOOPS05-BP02 Implement a structured sequence of interface assignment based on the persistent or ephemeral nature of the interfaces

The order of operations is important as the creation and termination of instances often creates and destroys interfaces. This is a useful property for ephemeral interfaces. However, it can lead to problems where other resources are dependent on the persistence of an interface and IP addressed across network domains. Persistent interfaces should be created prior to the creation of the EC2 instance with their properties assigned through IaC. Upon EC2 instance creation ephemeral interfaces and addresses can be created at the initialization of the EC2 instance while persistent interfaces which were previously created can be added to the EC2 instance. Thus, the termination of the instance will remove the ephemeral interfaces while preserving the persistent interfaces. The persistent interfaces retain their properties and can be re-assigned to a new EC2 instance without impacting adjacent functions.

Desired outcome:

- Clear interface lifecycle management.
- Predictable interface behavior.
- Proper resource dependency handling.
- Efficient interface assignment.
- Minimized service disruption.
- Reliable state management.

Common anti-patterns:

- Random interface assignment.
- No lifecycle consideration.
- Missing dependency mapping.
- Improper sequence handling.

- Poor state management.
- Undefined persistence rules.
- Inconsistent cleanup.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Create a structured approach to interface assignment that clearly distinguishes between persistent and ephemeral interfaces throughout their lifecycle. Develop automated workflows that handle the creation, attachment, and cleanup of interfaces in the correct sequence, making sure that persistent interfaces are properly preserved during instance replacements or updates. Implement state tracking mechanisms that maintain visibility into interface status and dependencies, enabling proper handling of complex networking scenarios. Establish robust error handling and rollback procedures to maintain system integrity during interface operations, while verifying proper cleanup of ephemeral resources.

Implementation steps

- Use AWS tags for interface type identification and AWS Resource Groups for interface categorization.
- Deploy AWS Step Functions for orchestrating interface assignment workflow and AWS Lambda for execution logic.
- Implement AWS CloudFormation for defining interface dependencies and AWS Systems Manager Automation for sequence execution.
- Use Amazon DynamoDB for interface state tracking and Amazon EventBridge for state change management.
- Configure AWS Systems Manager OpsCenter for interface-related operations and AWS CloudTrail for interface activity tracking.

Resources

Key AWS services:

- AWS CloudFormation
- AWS Systems Manager

- AWS Lambda
- Amazon EventBridge
- AWS Step Functions

Resource availability

TELCOOPS06: How do you verify resource availability for scale-out scenarios in telecommunication workloads?

Telecommunication workloads often require rapid scaling to meet sudden increases in demand or to maintain service continuity. However, cloud environments may not guarantee immediate availability of resources, especially for large instance types commonly used in telco deployments. This question addresses the challenges of maintaining consistent resource availability for scale-out scenarios in dynamic telecommunication environments.

Best practices

- TELCOOPS06-BP01 Verify resource availability for scale-out scenarios in telecommunication workloads

TELCOOPS06-BP01 Verify resource availability for scale-out scenarios in telecommunication workloads

If the cloud providers cannot guarantee instant EC2 instance availability, especially for large instance types commonly used in telco deployments, implement proactive capacity planning. Capacity planning can be broken into two classes: *new service* and *run-rate*.

New service planning requires cloud resources where there is no trend data to indicate a future demand profile, and the service team has no visibility to the future demand profile.

Run-rate planning indicates a stable environment where organic growth can be modeled and forecast based on service consumption.

Introduction of new services in an existing deployment requires both approaches as the composite demand is both run-rate and new service. The methods include utilizing capacity reservations,

maintaining buffer capacity, and implementing predictive scaling strategies based on historical usage patterns. Organizations should also consider using a mix of instance types and sizes that can support their workloads while increasing the likelihood of resource availability during scale-out events.

Desired outcome:

- Guaranteed resource availability.
- Efficient capacity planning.
- Predictable scaling operations.
- Optimized resource utilization.
- Minimized deployment delays.
- Cost-effective scaling.

Common anti-patterns:

- Reactive capacity planning.
- No resource forecasting.
- Missing buffer capacity.
- Poor utilization tracking.
- Inadequate reservation strategy.
- No scaling limits.
- Unplanned resource constraints.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement a comprehensive capacity management strategy that addresses both new service launches and ongoing operational needs. Create predictive scaling models that combine historical usage patterns with forecasted demand to verify adequate resource availability without over-provisioning. Develop buffer capacity policies that maintain appropriate headroom for unexpected demand spikes while considering cost implications and resource constraints. Establish regular capacity review processes that evaluate resource utilization trends, identify potential bottlenecks, and adjust capacity plans to maintain optimal performance and cost efficiency.

Implementation steps

- Use Service Quotas for limit management and AWS Compute Optimizer for resource optimization recommendations.
- Implement AWS Auto Scaling for dynamic resource adjustment and EC2 Fleet for capacity diversity.
- Deploy AWS Systems Manager Automation for scaling procedures and Amazon EventBridge for automated scaling triggers.
- Configure AWS Cost Explorer for resource cost tracking and AWS Budgets for cost management.
- Use AWS Systems Manager OpsCenter for capacity management and Amazon CloudWatch for utilization monitoring.

Resources

Key AWS services:

- AWS Auto Scaling
- Amazon EC2 Fleet
- AWS Capacity Manager
- AWS Systems Manager
- AWS Application Auto Scaling
- AWS Service Quotas

Geographical redundancy

TELCOOPS07: How do you address geographical redundancy requirements within data residency constraints?

Telecommunication service providers often face strict requirements for geographical redundancy (for example, 200 or more kilometers of separation) while simultaneously needing to comply with data residency regulations. This creates unique challenges in cloud environments where multiple regions within a single country may not be available. This question explores strategies

for balancing redundancy needs with regulatory adherence in cloud-based telecommunication architectures.

Best practices

- TELCOOPS07-BP01 Design architecture patterns that maximize redundancy within available geographical constraints

TELCOOPS07-BP01 Design architecture patterns that maximize redundancy within available geographical constraints

For telco CSPs requiring 200 or more kilometers of geo-redundancy while maintaining in-country data residency, it is essential to design cloud-based architectural solutions within available geographical constraints. This may include leveraging available Availability Zones, Local Zones, or Wavelength Zones where available, implementing hybrid architectures, or working with cloud providers to understand future infrastructure expansion plans. The solution should balance regulatory requirements with disaster recovery needs while maintaining required service levels.

Desired outcome:

- Optimal geographical redundancy.
- Adherence with data residency.
- Resilient architecture design.
- Efficient disaster recovery.
- Maintained service levels.
- Balanced resource distribution.

Common anti-patterns:

- Single region deployments.
- Insufficient separation distance.
- Non-compatible data placement.
- Poor failover planning.
- Missing redundancy testing.
- Inadequate backup strategies.
- Unclear recovery procedures.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Develop architectural solutions that optimize geographical redundancy while adhering to data residency requirements and regulatory constraints. Create detailed mapping of available infrastructure options, including primary locations, disaster recovery sites, and edge locations, to design optimal redundancy patterns that meet separation requirements. Implement robust data replication and synchronization mechanisms that maintain consistency across geographical locations while minimizing latency and bandwidth usage. Establish comprehensive failover procedures and regular testing protocols to validate the effectiveness of geographical redundancy implementations, verifying that service levels are maintained during both planned and unplanned events.

Implementation steps

- Use AWS Local Zones and Wavelength for edge presence and AWS Global Infrastructure for Region or zone selection based on regulatory requirements.
- Implement AWS Transit Gateway for inter-region connectivity and AWS Global Accelerator for intelligent traffic routing.
- Deploy AWS CloudFormation StackSets for multi-Region deployment and Amazon Route 53 for DNS-based failover.
- Use AWS Fault Injection Service for disaster recovery testing and AWS Systems Manager Automation for failover procedures.
- Configure AWS Network Manager for global network visibility and Amazon CloudWatch for cross-Region monitoring.

Resources

Key AWS services:

- AWS Local Zones
- AWS Wavelength
- Amazon Route 53
- AWS Global Accelerator
- AWS Transit Gateway

- AWS Site-to-Site VPN

Failure testing

TELCOOPS08: How do you implement comprehensive failure testing for telecommunication workloads in cloud environments?

Cloud environments present unique failure modes and behaviors that may differ significantly from on-premises infrastructure. Telecommunication workloads, given their critical nature, require robust failure testing to verify resilience. This question addresses the need for adapting traditional testing approaches to account for cloud-specific characteristics and implementing comprehensive failure testing programs for telecommunication services.

Best practices

- TELCOOPS08-BP01 Develop and execute comprehensive failure testing programs that reflect cloud infrastructure characteristics

TELCOOPS08-BP01 Develop and execute comprehensive failure testing programs that reflect cloud infrastructure characteristics

ISVs and CSPs should adapt their testing approaches to account for cloud infrastructure characteristics that differ from on-premises environments. This includes developing comprehensive failure testing programs that simulate various network and infrastructure failures, implementing chaos engineering practices, and using tools like AWS Fault Injection Service to test system resilience. Tests should go beyond basic instance or pod failures to include network impairments, Availability Zone failures, and other realistic failure scenarios that could impact telecommunication services.

Desired outcome:

- Comprehensive failure testing.
- Validated resilience measures.
- Identified weaknesses.
- Improved recovery procedures.

- Verified failover capabilities.
- Enhanced system reliability.

Comment anti-patterns:

- Limited test scenarios
- Unrealistic test conditions
- Missing failure scenarios
- Poor test documentation
- Inadequate recovery testing
- No regular testing schedule
- Incomplete impact analysis

Level of risk exposed if this best practice is not established: High

Implementation guidance

Design and implement a systematic failure testing program that validates system resilience across critical components and services. Create a comprehensive test matrix that includes various failure scenarios, from individual component failures to complete zone or regional outages, while considering cloud-specific characteristics and behaviors. Implement automated testing frameworks that can safely simulate failures and validate recovery procedures without risking production workloads. Establish detailed monitoring and analysis procedures to capture test results, identify improvement areas, and track the evolution of system reliability over time.

Implementation steps

- Use AWS Fault Injection Service for controlled chaos experiments and AWS Systems Manager for test automation.
- Deploy AWS Step Functions for orchestrating test scenarios an d AWS Lambda for test execution.
- Implement Amazon EventBridge for test scheduling and AWS Systems Manager Automation for test procedures.
- Use Amazon CloudWatch for test result collection and Amazon OpenSearch Service for test data analysis.
- Configure AWS Systems Manager OpsCenter for test management and AWS CloudTrail for test activity logging.

Resources

Key AWS services:

- AWS Fault Injection Service (FIS)
- Amazon CloudWatch
- AWS Systems Manager
- AWS Lambda
- AWS Step Functions
- AWS Health Dashboard

Performance baselining

TELCOOPS09: How do you verify accurate performance baseline measurements for capacity planning?

In cloud environments, particularly those with shared underlying infrastructure, accurately measuring performance baselines for telecommunication workloads can be challenging. The lack of visibility into resource utilization at the hardware level and potential variations in performance due to multi-tenancy can lead to inaccurate capacity planning. This question explores strategies for establishing reliable performance baselines in cloud environments to support effective capacity planning for telecommunication workloads.

Best practices

- TELCOOPS09-BP01 Implement comprehensive performance monitoring and baseline testing that accounts for shared infrastructure characteristics

TELCOOPS09-BP01 Implement comprehensive performance monitoring and baseline testing that accounts for shared infrastructure characteristics

To address the challenges of performance baselining in cloud environments with shared resources (like Nitro-based instances), organizations should implement comprehensive monitoring and

testing strategies that account for various load conditions. This includes monitoring instance-level metrics, understanding resource allocation limits, and conducting performance tests under different multi-tenant scenarios. Organizations should also leverage available performance metrics and tools to understand resource utilization patterns and verify accurate capacity planning that accounts for shared infrastructure characteristics.

Desired outcome:

- Accurate performance baselines.
- Reliable capacity planning.
- Comprehensive monitoring.
- Performance predictability.
- Resource optimization.
- Early issue detection.

Common anti-patterns:

- Inconsistent performance metrics.
- Missing baseline data.
- Poor monitoring coverage.
- Inadequate testing scenarios.
- No consideration of multi-tenancy.
- Unrealistic benchmarks.
- Lack of historical analysis.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Establish a holistic performance monitoring framework that accounts for the unique characteristics of shared cloud infrastructure and telecommunication workloads. Develop comprehensive baseline testing procedures that capture performance metrics across different load conditions, time periods, and infrastructure configurations to create reliable performance profiles. Implement continuous monitoring solutions that track key performance indicators while accounting for multi-tenant impacts and resource contention scenarios. Create analysis frameworks that combine historical

trends, real-time metrics, and predictive analytics to enable proactive performance optimization and capacity planning.

Implementation steps

- Deploy Amazon CloudWatch for metric collection and Amazon Managed Grafana for visualization.
- Configure AWS X-Ray for distributed tracing and Amazon Managed Service for Prometheus for metric storage.
- Use Amazon CloudWatch Synthetics for performance testing and AWS Lambda for custom test execution.
- Implement Amazon OpenSearch Service for log analytics and Quick for performance analytics.
- Deploy AWS Systems Manager OpsCenter for performance issue management and AWS Compute Optimizer for resource optimization.

Resources

Key AWS services:

- Amazon CloudWatch
- AWS X-Ray
- Amazon Managed Grafana
- Amazon Managed Service for Prometheus
- AWS CloudTrail
- Amazon OpenSearch Service

Security

As communications service providers (CSPs) migrate their workloads to the public cloud, it is crucial that they implement robust security measures to protect their infrastructure and data. Recent cybersecurity incidents have highlighted the need for enhanced security visibility and hardening of communications systems. The public cloud, with its secure-by-design and secure-by-default philosophy, can provide CSPs with the necessary tools and capabilities to effectively address this guidance. By understanding the shared responsibility model and leveraging the scale and automation advantages of the public cloud, CSPs can strengthen their security posture, increase visibility, and implement rigorous access controls, cryptography, and vulnerability management practices. Embracing these security best practices is essential for CSPs to safeguard their critical communications infrastructure in the cloud environment.

Focus areas

- Design principles
- Security foundations
- Infrastructure protection
- External interface security
- Data Protection

Design principles

- **Secure by design and secure by default:** Public cloud services are designed with security as a core principle. Cloud resources are secure by default, with features like isolated network environments, private storage, and encrypted network traffic enabled out-of-the-box. This secure-by-design approach verifies that customers benefit from robust security measures without the need for extensive manual configuration, reducing the risk of misconfigurations and human errors.
- **Shared responsibility model:** The shared responsibility model in the public cloud clearly delineates security responsibilities between the cloud provider and the customer. The provider is responsible for the security of the underlying cloud infrastructure, while customers are responsible for the security of their applications and data within the cloud. This division of responsibilities significantly reduces the operational burden on customers, allowing them to

focus on protecting their workloads while the provider handles the foundational security of the cloud.

- **Automation and scalability:** The massive scale of public cloud operations enables extensive automation of security tasks, such as feature and patch deployments, configuration updates, and threat mitigation. This automation reduces the risk of human errors and provides customers with the benefits of continuously updated security measures. Furthermore, the scale of public cloud providers allows for comprehensive threat intelligence and adherence programs, which customers can leverage to enhance their overall security posture.
- **Workload isolation and visibility**: The public cloud's multi-tenant architecture, with strong tenant isolation, provides customers with inherent visibility and control over their resources. In the public cloud, resources are isolated by default, and explicit configuration is required to enable access and connectivity between them. This isolation, combined with comprehensive logging and monitoring capabilities, empowers customers to maintain a clear understanding of their digital assets and network activities, which is crucial for detecting and responding to potential threats.

Security foundations

TELCOSEC01: Is your network compatible with the telecom industry and regulatory requirements?

Telecom networks must comply with industry specifications and local regulatory requirements to demonstrate a commitment to security and build customer trust. This question focuses on how the organization verifies its network and operations adhere to the relevant mandates in its serving regions.

Best practices

- TELCOSEC01-BP01 Follow the industry standards and local regulations to host telecom networks in the cloud, demonstrating a security commitment
- TELCOSEC01-BP02 Encrypt data at-rest and sensitive traffic (control plane, data plane) using secure VPNs and custom encryption keys

TELCOSEC01-BP01 Follow the industry standards and local regulations to host telecom networks in the cloud, demonstrating a security commitment

When operating a telecommunications network in a cloud environment, it is crucial to comply with relevant industry specifications and local regulatory requirements. This includes adhering to standards set forth by leading telecom standardization bodies such as the 3rd Generation Partnership Project (3GPP), which develops technical specifications for mobile networks, and the European Telecommunications Standards Institute (ETSI), which produces globally applicable standards for information and communications technologies.

Additionally, the network operations must meet the regulations enforced by authorities like the Federal Communications Commission (FCC) in the United States and the European Union's Electronic Communications Code. By aligning with these industry specifications and local laws, the organization demonstrates a strong commitment to security and building trust with customers.

Adhering to established standards and regulations verifies the cloud-based telecommunications network functions reliably and safeguards sensitive customer data. This approach fosters lasting relationships with the customer base through transparent, secure, and well-regulated operations.

Implementation guidance

Establish a comprehensive security hardening strategy that protects your AWS infrastructure through multi-layered defense mechanisms and continuous monitoring. Design a robust security architecture that encompasses identity management, network segmentation, encryption protocols, and vulnerability management while maintaining operational efficiency. Implement centralized logging and monitoring capabilities that provide visibility across the layers of your infrastructure, enabling rapid detection and response to security incidents. Create detailed security policies and operational procedures that enforce least privilege access, automate checks, and maintain audit trails for configuration changes to verify the integrity and resilience of your cloud environment.

Implementation steps

- Use Amazon Security Lake for centralized logging.
- Deploy AWS Config for configuration tracking.
- Implement Amazon CloudWatch for monitoring and alerting.
- Enable AWS CloudTrail for API activity logging.
- Configure AWS IAM Identity Center for authentication.

- Set up AWS IAM for role-based access control.
- Deploy AWS Secrets Manager for credential management.
- Enable MFA for privileged accounts.
- Design Amazon VPC architecture for network isolation.
- Configure AWS Transit Gateway for traffic management.
- Implement AWS Network Firewall for traffic filtering.
- Set up security groups and NACLs for micro-segmentation.
- Deploy AWS KMS for key management.
- Configure AWS Certificate Manager for TLS certificates.
- Implement AWS CloudHSM for hardware security modules.
- Enable encryption at rest for data stores.
- Implement AWS Systems Manager for patch management.
- Deploy Amazon Inspector for vulnerability scanning.
- Configure AWS Security Hub CSPM for security posture monitoring.
- Establish automated remediation workflows.

For more detail, see Enhancing telecom security with AWS.

Resources

Related documents:

- Enhancing telecom security with AWS

Key AWS services:

- AWS Key Management Service (KMS)
- Amazon VPC
- AWS IAM
- AWS CloudHSM
- AWS CloudTrail
- AWS Config
- AWS Systems Manager

TELCOSEC01-BP02 Encrypt data at-rest and sensitive traffic (control plane, data plane) using secure VPNs and custom encryption keys

Due to regulatory requirements, the CSPs often need to apply additional encryption to traffic that is not already secured by the underlying transport or application protocols. Both control plane traffic, such as Diameter signaling, and data plane traffic, like voice or RTP, require the use of encrypted virtual private networks (VPNs), most commonly IPsec. It is crucial for CSPs to evaluate their regulatory landscape and establish a framework to identify and label the traffic and data that needs to be encrypted in transit as well as at rest, using customer-owned encryption keys.

The data in-use is protected by the underlying CPU architecture and features like AMD SEV-SNP, Intel TME, and Graviton Memory Encryption. However, these security measures come with additional costs and limitations, such as restrictions on packets per second (PPS) per flow (number of IPsec tunnels), PPS per instance (lack of equal-cost multi-path routing after reaching the largest single-slot capacity), complexity, and vendor dependencies.

Desired outcome:

- Verify sensitive telco network traffic, including control plane and data plane, is encrypted using secure VPN protocols like IPsec.
- Implement a comprehensive data encryption strategy to protect data in transit and at rest using customer-managed encryption keys.
- Fulfill regulatory requirements for data protection and secure communications in the telco industry.
- Maintain the confidentiality and integrity of critical telco network data and signaling traffic, even when traversing the public cloud infrastructure.
- Provide visibility and control over the encryption keys and algorithms used to secure the telco workloads.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement a secure, encrypted communication framework for control and data plane traffic comply with industry standards and regulations. Leverage built-in features for key management, high availability, and automatic failover. Utilize a centralized key management service to generate, manage, and rotate encryption keys for securing data in-transit and at-rest across various systems

and services. Verify comprehensive auditing and monitoring of encryption and key management activities to maintain adherence and provide detailed reporting.

Implementation steps

- Implement VPN-based encryption for control and data plane traffic:
  - Configure AWS managed VPN solutions, such as Site-to-Site VPN, to establish secure IPsec tunnels for control and data plane communications.
  - Use the built-in features of these services for automatic key management, tunnel failover, and high availability.
  - Verify the VPN configurations comply with industry standards and regulations for secure network communications.
- Encrypt data in-transit using custom keys:
  - Use AWS Key Management Service to generate and manage the encryption keys used to secure data in-transit.
  - Integrate KMS with other AWS services, such as Amazon S3, Amazon EBS, and Amazon RDS, to transparently encrypt data flowing through these services.
  - Implement a key rotation strategy to maintain the cryptographic strength of the encryption keys over time.
- Encrypt data at rest using custom keys:
  - Configure AWS services handling telco data, such as Amazon S3, Amazon EBS, and Amazon RDS, to use customer-managed encryption keys from KMS.
  - Verify that sensitive telco data, including logs, metrics, and configuration files, is encrypted at rest using the custom encryption keys.
  - Use KMS key policies to control access and usage of the encryption keys, aligned with the principle of least privilege.
- Implement key management and rotation:
  - Establish key management processes and procedures for the creation, distribution, and rotation of the customer-managed encryption keys.
  - Automate key rotation using AWS Lambda functions or AWS Systems Manager, verifying that keys are updated on a regular basis.
  - Integrate the key management processes with the overall security operations and incident response procedures.
- Maintain audit trails:

- Use AWS CloudTrail to create comprehensive audit trails of key management and data encryption activities.
- Configure AWS Config to continuously monitor the configuration of the encryption services and report on deviations from the desired state.
- Integrate the encryption and key management data with the organization's security information and event management (SIEM) system for advanced analytics and reporting.

Resources

Key AWS services:

- Site-to-Site VPN
- AWS Direct Connect
- AWS Key Management Service (KMS)
- Amazon S3
- Amazon EBS
- Amazon RDS
- AWS CloudTrail
- AWS Config
- AWS Lambda
- AWS Systems Manager

Infrastructure protection

TELCOSEC02: How do you implement holistic security within your telco network to avoid risks and provide integrity?

Telecom networks face diverse security threats, making it crucial for operators to implement comprehensive end-to-end security measures to block unauthorized access, malicious attacks, and data breaches while providing for confidentiality, integrity, and availability of your network services and subscriber information.

Best practices

- TELCOSEC02-BP01 Use secure control plane protocols between the network functions (NF)
- TELCOSEC02-BP02 Implement features such as concealed or spoofed subscriber identities for signaling over the interfaces with access network
- TELCOSEC02-BP03 Segment and isolate telco network domains and slices

TELCOSEC02-BP01 Use secure control plane protocols between the network functions (NF)

The 3rd Generation Partnership Project (3GPP) recommends that telco organizations use secure control plane protocols between the various network functions (NF) within their 5G core network architecture. This best practice is crucial for maintaining the confidentiality, integrity, and availability of the control plane communications, which are essential for the proper functioning and management of the network.

The control plane is responsible for the exchange of signaling and management information between network elements, handling critical tasks such as session establishment, mobility management, and policy enforcement. By implementing secure control plane protocols, telco organizations can mitigate the risk of unauthorized access, eavesdropping, and tampering of the control plane traffic, thereby enhancing the overall security and resilience of the network.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement secure communication protocols, such as HTTPS and TLS/DTLS, for control plane traffic to verify encryption and integrity protection. Use a trusted and secure mechanism for network function registration and discovery, integrating it with a central registry to maintain an authorized list of network components. Monitor and analyze the control plane traffic for anomalies using comprehensive logging and advanced analytics and develop robust incident response procedures to address security incidents related to the control plane, with a focus on automated workflows and rapid remediation.

Implementation steps

Implement secure control plane protocols:

- Verify that the control plane protocols are configured to use secure protocol such as HTTPs and secure communication channels, such as TLS/DTLS for encryption and integrity protection.
- Use AWS Certificate Manager to manage and provision the necessary SSL/TLS certificates for the control plane protocol endpoints.

Secure NF registration and discovery:

- Use a secure NF registration and discovery mechanism, such as the ones defined in 3GPP TS 29.510, to enable trusted NFs to discover and communicate with each other.
- Integrate the NF registration and discovery process with AWS Cloud Directory or Service Catalog to maintain a secure and up-to-date registry of authorized network functions.

Control plane traffic monitoring and anomaly detection:

- Implement comprehensive logging and monitoring of the control plane traffic using Amazon CloudWatch and Amazon OpenSearch Service.
- Configure Amazon GuardDuty to detect and alert on anomalous or suspicious activities within the control plane protocol communications.
- Use Amazon CloudWatch Logs Insights to perform advanced analysis and forensics on the control plane traffic logs.

Control plane incident response and remediation:

- Develop and test incident response procedures to address security incidents or breaches related to the control plane protocols.
- Integrate the control plane security monitoring and incident response capabilities with the overall security operations and incident management processes.
- Use AWS Lambda, AWS Step Functions, and Amazon SNS to automate the incident response workflows and facilitate rapid remediation.

Resources

Key AWS services:

- AWS Lambda

- AWS Certificate Manager
- AWS Cloud Directory
- Service Catalog
- Amazon OpenSearch Service
- AWS Step Functions
- Amazon CloudWatch
- Amazon GuardDuty
- Amazon SNS

TELCOSEC02-BP02 Implement features such as concealed or spoofed subscriber identities for signaling over the interfaces with access network

The 3rd Generation Partnership Project (3GPP) recommends that telco organizations use the Subscription Concealed Identifier (SUCI) instead of the Subscription Permanent Identifier (SUPI) on the air interface between the user equipment (UE) and the access network. This best practice is aimed at enhancing the privacy and security of subscriber identities, as the air interface is considered the most exposed and vulnerable part of the network.

The SUPI, which includes sensitive subscriber information such as the International Mobile Subscriber Identity (IMSI), is a unique and permanent identifier associated with each subscriber. By using the SUCI, a privacy-preserving temporary identifier derived from the SUPI, telco organizations can block the disclosure of the actual subscriber identity over the air interface, mitigating the risk of subscriber tracking, profiling, and other privacy-related attacks.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement a secure mechanism for generating and processing the Subscription Concealed Identifier (SUCI) instead of the Subscription Permanent Identifier (SUPI) for signaling over the air interface, maintaining adherence with 3GPP privacy protection schemes. Establish a secure system for mapping and managing the SUPI-SUCI relationship, with robust encryption and integrity protection for the air interface communications. Monitor the air interface signaling for anomalies and develop incident response procedures to address security incidents related to subscriber

identity, prioritizing privacy protection. Verify comprehensive security governance and adherence with industry standards and regulations for the air interface security controls.

Implementation steps

Implement SUCI generation and conversion:

- Configure the UE and the Access and Mobility Management Function (AMF) to generate and process the SUCI instead of the SUPI for signaling over the air interface.
- Verify that the SUCI is generated using the appropriate privacy protection scheme, as defined in 3GPP TS 33.501.

Incident response and subscriber identity protection:

- Develop and test incident response procedures to address security incidents or breaches related to the SUCI signaling over the air interface.
- Use AWS Lambda, AWS Step Functions, and Amazon SNS to automate the incident response workflows and facilitate rapid remediation.
- Verify that the incident response plan includes steps for investigating, mitigating, and reporting subscriber identity-related security incidents, while preserving the privacy of the affected subscribers.

Air interface security governance:

- Establish security governance policies and controls for the management of SUCI signaling over the air interface.
- Use AWS Config, AWS Security Hub CSPM, and AWS Trusted Advisor to continuously assess the security posture and verify adherence with industry standards and regulations, such as GDPR.

Resources

Key AWS services:

- AWS Key Management Service (KMS)
- AWS Certificate Manager
- Amazon CloudWatch

- Amazon GuardDuty
- AWS Step Functions
- Amazon SNS

TELCOSEC02-BP03 Segment and isolate telco network domains and slices

To maintain the security and resilience of a telco network in a cloud environment, segment and isolate the various domains within the network. The same understanding applies to the situation where multiple virtual network slices are hosted in a common environment. This approach assists to mitigate the risk of potential security breaches and minimize the impact of an incident.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Establish logical isolation by creating distinct network environments for the various functional domains within the telco network, such as control plane, user plane, management plane, and signaling plane. Implement robust access controls and monitoring mechanisms to restrict and govern cross-domain communication. Achieve physical isolation by leveraging geographically separated infrastructure, if required, and enforce strict access control and identity management. Facilitate secure connectivity between the isolated network domains through encrypted and integrity-protected communication channels and manage the necessary certificates and keys centrally.

Implementation steps

Logical isolation:

- Identify the distinct functional domains within the telco network, such as control plane, user plane, management plane and signaling plane.
- Use AWS Virtual Private Cloud (VPC) to create distinct and isolated network environments for the different functional domains within the telco network, such as control plane, user plane, management plane, and signaling plane.
- Implement network access controls using AWS security groups and network access control lists (NACLs) to restrict and monitor cross-domain communications.
- Use AWS VPC Peering to enable secure connectivity between the isolated VPCs, if required.

Physical isolation:

- Use AWS Availability Zones and Regions to physically separate the network domains, if needed.
- Use AWS Outposts or AWS Local Zones to deploy dedicated infrastructure for specific network domains, maintaining physical isolation.
- Integrate AWS Identity and Access Management to enforce strict access control and identity management for personnel and systems interacting with the isolated domains.

Secure connectivity:

- Establish secure communication using AWS Direct Connect or Site-to-Site VPN for interconnecting the isolated network domains.
- Use AWS Certificate Manager to manage and renew SSL/TLS certificates for secure communication.

Resources

Key AWS services:

- AWS VPC
- AWS Security Groups
- AWS Availability Zones
- AWS Local Zones
- AWS Outposts
- AWS Identity and Access Management (IAM)
- AWS Direct Connect
- Site-to-Site VPN
- AWS Private Link
- AWS PrivateSubnet
- AWS Certificate Manager

External interface security

TELCOSEC03: How do you maintain security on the external interfaces of your telco network?

Telecom networks expose critical functionality through external interfaces that include APIs, control plane protocols, and data traffic. Maintaining robust security across these access points is essential to block unauthorized access and protect the network's integrity. This question explores how the organization implements security controls and updates mitigation measures to verify only authorized systems can interact with the network's external touchpoints.

Best practices

- TELCOSEC03-BP01 Secure the APIs used to expose telco network functionality
- TELSEC03-BP02 Deploy signaling firewall on the roaming and interconnecting interfaces with other telco networks
- TELSEC03-BP03 Perform regular penetration testing on each of the protocols implemented on the signaling layer

TELCOSEC03-BP01 Secure the APIs used to expose telco network functionality

In modern telco networks, the Network Exposure Function (NEF) for 5G networks and the Service Capability Exposure Function (SCEF) for 4G networks provide APIs to expose network capabilities and services to external applications and partners. These APIs serve as the gateway for accessing sensitive network resources and data, making it crucial to implement robust security measures to protect them.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implement strong authentication mechanisms and enforce role-based access control to authorize API access based on client permissions and privileges. Manage user identities and generate secure JSON web tokens. Verify API communication is encrypted using HTTPS/TLS protocols and leverage web application firewalls to enforce and monitor security. Implement API rate limiting and

throttling to block potential denial-of-service attacks. Perform input validation and sanitization on API requests and integrate comprehensive logging and monitoring to detect and respond to suspicious API activity. Manage API versioning and deprecation to verify clients use the latest secure versions and regularly assess API security through automated testing and vulnerability management.

Implementation steps

- API traffic encryption:
  - Verify API communication is encrypted using HTTPS/TLS protocols with AWS Certificate Manager to manage SSL/TLS certificates.
  - Leverage AWS WAF (Web Application Firewall) to enforce SSL/TLS adherence and monitor TLS cipher suite usage.
- API rate limiting and throttling:
  - Implement API rate limiting and throttling using Amazon API Gateway's built-in throttling capabilities.
  - Configure API Gateway to enforce rate limits and block potential denial-of-service (DoS) attacks.
- API input validation and sanitization:
  - Utilize AWS Lambda functions as API backends to perform input validation and sanitization using libraries like express-validator or Joi.
  - Integrate Amazon API Gateway with AWS WAF to apply custom security rules for input validation and protection against common web application vulnerabilities.
- API logging and monitoring:
  - Enable comprehensive logging of API requests and responses using Amazon CloudWatch Logs.
  - Set up Amazon CloudWatch alarms and Amazon SNS notifications to alert on suspicious API activity.
- API versioning and deprecation:
  - USe Amazon API Gateway's versioning capabilities to manage multiple versions of the NEF and SCEF APIs.
  - Implement version-specific access controls using AWS IAM policies to verify clients use the latest secure API versions.
  - Utilize Amazon API Gateway's stage management features to deprecate older API versions and provide clear migration guidelines.

- API security assessments:
  - Perform API security assessments using AWS Security Hub CSPM, which aggregates findings from various AWS security services.
  - Address identified vulnerabilities using AWS Systems Manager for patch management and AWS Lambda for rapid remediation.

Resources

Key AWS services:

- AWS Identity and Access Management (IAM)
- AWS Certificate Manager
- AWS WAF
- Amazon API Gateway
- AWS Lambda
- Amazon CloudWatch
- AWS Security Hub CSPM
- AWS Systems Manager

TELSEC03-BP02 Deploy signaling firewall on the roaming and interconnecting interfaces with other telco networks

The Global System for Mobile Communications Association (GSMA) recommends deploying signaling firewalls on the roaming and interconnecting interfaces with other telco networks. This best practice is aimed at protecting the telco network from potential signaling-based attacks and maintaining the overall security and integrity of the network.

The signaling firewall acts as a gatekeeper, monitoring and filtering the signaling traffic exchanged between the telco network and its roaming partners or interconnected networks. By implementing a signaling firewall, organizations can effectively mitigate the risk of unauthorized access, signaling storms, and other signaling-related security threats, thereby enhancing the resilience and reliability of the network.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Catalog and understand the network topology and signaling traffic across each identified interface. Deploy a signaling firewall solution capable of deep packet inspection, protocol validation, and anomaly detection, aligned with the organization's security policies. Integrate the signaling firewall with comprehensive logging, monitoring, and centralized security analysis capabilities. Configure the firewall to inspect and filter signaling traffic, implementing protocol-specific rules to detect and block unauthorized or malicious activities. Establish robust monitoring and incident response procedures, integrating the firewall logs with advanced security analytics tools. Regularly review and update the signaling firewall's configuration, rulesets, and software versions to maintain effectiveness against evolving threats, leveraging automated maintenance and validation processes.

Implementation steps

- Identify roaming and interconnection interfaces:
  - Catalog the roaming and interconnection interfaces within the telco network, including the protocols and signaling traffic exchanged.
  - Understand the network topology and the flow of signaling traffic across the identified interfaces.
- Deploy signaling firewall solution:
  - Select a signaling firewall solution that is compatible with the identified protocols and interfaces within the telco network.
  - Configure the signaling firewall to align with the organization's security policies and best practices.
- Integrate signaling firewall with AWS services:
  - Integrate the signaling firewall with Amazon CloudWatch for comprehensive logging and monitoring of signaling traffic and security events.
  - Utilize AWS Security Hub CSPM to centralize the security findings from the signaling firewall and other security services, enabling holistic security analysis and incident response.
- Signaling traffic inspection and filtering:
  - Configure the signaling firewall to inspect incoming and outgoing signaling traffic, validating the integrity and authenticity of the signaling messages.

  - Implement signaling protocol-specific rules and policies to detect and block unauthorized or malicious signaling activities, such as signaling storms, fraud attempts, and network element impersonation.
  - Regularly update the signaling firewall's rulesets and signatures to address evolving signaling-based threats and vulnerabilities.
- Signaling firewall monitoring and incident response:
  - Establish robust monitoring and alerting mechanisms to detect and respond to anomalies or security incidents related to the signaling traffic.
  - Integrate the signaling firewall logs with Amazon CloudWatch and Amazon OpenSearch Service for advanced security analysis and threat hunting.
  - Develop and test incident response procedures to mitigate the impact of signaling-based attacks or network disruptions.
- Signaling firewall maintenance and updates:
  - Regularly review and update the signaling firewall's configuration, rulesets, and software versions to verify it remains effective against the latest security threats.
  - Leverage AWS Systems Manager for automated patching and updates of the signaling firewall infrastructure.
  - Conduct periodic testing and validation of the signaling firewall's functionality and security posture.

Resources

Key AWS services:

- Amazon OpenSearch Service
- Amazon CloudWatch
- AWS Security Hub CSPM
- AWS Systems Manager

TELSEC03-BP03 Perform regular penetration testing on each of the protocols implemented on the signaling layer

Perform regular penetration testing on each of the protocols implemented on the signaling layer of their network. This best practice is crucial for identifying and addressing vulnerabilities that

could be exploited by malicious actors to compromise the security and integrity of the telco network.

The signaling layer is responsible for the exchange of control and management information between network elements, making it a critical attack surface. Penetration testing on the signaling layer protocols assists organizations to uncover potential weaknesses, such as improper protocol implementation, cryptographic flaws, or insecure configurations, which could lead to unauthorized access, data breaches, or service disruptions.

By conducting regular and comprehensive penetration testing, telco organizations can proactively identify and mitigate risks, maintaining the overall resilience and security of their signaling infrastructure.

Implementation guidance

Maintain a comprehensive inventory of implemented signaling layer protocols, including their versions and implementation details. Define the scope and objectives of a penetration testing exercise focused on the identified signaling protocols, developing a detailed testing plan aligned with security policies and regulatory requirements. Conduct thorough testing of the signaling protocols, verifying the proper implementation of security controls, and identifying vulnerabilities or misconfigurations. Prioritize the discovered vulnerabilities based on their risk and potential impact, and develop and execute remediation plans to address them, leveraging automated patch management and security update deployment processes.

Implementation steps

- Signaling layer protocol inventory:
  - Identify the signaling layer protocols implemented within the telco network, such as Diameter, SS7, SIP, and GTP.
  - Maintain a comprehensive inventory of the signaling protocols, including their versions and implementation details.
- Penetration testing scoping and planning:
  - Define the scope and objectives of the penetration testing exercise, focusing on the identified signaling layer protocols.
  - Verify the penetration testing activities are aligned with the organization's security policies and regulatory requirements.
- Use AWS security services for penetration testing:

  - Utilize AWS Security Hub CSPM to centralize the findings and recommendations from the penetration testing activities.
  - Leverage AWS Systems Manager for efficient and secure deployment of the penetration testing tools and environments.
- Comprehensive signaling protocol testing:
  - Conduct thorough testing of the identified signaling layer protocols, including fuzzing, protocol-specific vulnerability scanning, and exploitation attempts.
  - Verify the proper implementation of security controls, such as encryption, authentication, and authorization, within the signaling protocols.
  - Identify and document vulnerabilities or misconfigurations that could be exploited to compromise the signaling layer.
- Remediation and vulnerability management:
  - Prioritize the identified vulnerabilities based on their risk and potential impact on the telco network.
  - Develop and execute remediation plans to address the discovered vulnerabilities, including software updates, configuration changes, and the implementation of additional security controls.
  - Use AWS Systems Manager for efficient patch management and deployment of security updates across the signaling layer infrastructure.

Resources

Key AWS services:

- Amazon CloudWatch
- AWS Security Hub CSPM
- AWS Systems Manager

Data Protection

TELCOSEC04: How do you protect customer proprietary network information (CPNI) and personally identifiable information (PII)?

Telecom operators must protect sensitive customer data like CPNI and PII in accordance with strict regulations. This question examines the organization's implementation of encryption, access controls, audit logging, and data handling procedures to block unauthorized access or misuse of customer information.

Best practices

- TELCOSEC04-BP01 Enable encryption for CPNI and PII information at rest and in transit and access restriction

TELCOSEC04-BP01 Enable encryption for CPNI and PII information at rest and in transit and access restriction

It is recommended that telecommunications organizations implement comprehensive encryption measures for both customer proprietary network information (CPNI) and personally identifiable information (PII) at rest and in transit to verify robust data protection and regulatory adherence. Organizations should use industry-standard encryption algorithms (such as AES-256) for data at rest, implement TLS/SSL protocols for data in transit, and establish secure key management practices.

This implementation should include encryption where applicable, strong access controls with multi-factor authentication, and regular security audits to verify encryption effectiveness. The organization must verify that third-party vendors handling CPNI or PII adhere to the same encryption standards and maintain detailed documentation of encryption practices.

Furthermore, establish comprehensive employee training programs to verify proper handling of encrypted data, and incident response plans should be developed specifically for potential breaches of encrypted information. This recommendation is critical for avoiding potential data breaches, maintaining regulatory adherence, protecting against financial penalties, and preserving customer trust. Conduct regular reviews and updates of encryption protocols to address emerging security threats and maintain the effectiveness of data protection measures.

Desired outcome:

- Verify that customer proprietary network information (CPNI) and personally identifiable information (PII) data is encrypted at rest and in transit, providing a robust layer of protection for sensitive customer data.
- Implement strict access controls and least-privilege permissions to limit access to CPNI and PII data, reducing the risk of unauthorized access or data breaches.

- Demonstrate adherence with industry regulations and customer data privacy requirements, such as FCC CPNI rules and GDPR, by effectively securing the handling of sensitive information.
- Provide visibility and auditability of access and usage of CPNI and PII data through comprehensive logging and monitoring.

Level of risk exposed if this best practice is not established: High

Implementation guidance:

Implement robust data protection measures by encrypting customer CPNI and PII data at rest and in transit. Use customer-controlled encryption keys and secure communication protocols like TLS. Enforce strict access controls based on the principle of least privilege, with multi-factor authentication. Enable comprehensive logging and monitoring to track access and usage activities, integrating the logs into a centralized security framework. Establish clear data handling policies and procedures and conduct regular audits to verify the effectiveness of the data protection controls and verify ongoing adherence with industry regulations and customer privacy requirements.

Implementation steps

- Encrypt CPNI and PII data at rest:
  - Use AWS Key Management Service (KMS) to generate and manage customer-controlled encryption keys for protecting data at rest.
  - Configure AWS services handling CPNI and PII data, such as Amazon S3, Amazon RDS, and Amazon EBS, to use the customer-managed encryption keys from KMS.
  - Implement a key rotation strategy to verify the cryptographic strength of the encryption keys is maintained over time.
- Encrypt CPNI and PII data in transit:
  - Require the use of secure communication protocols, such as TLS 1.2 or 1.3, for data transfers involving CPNI and PII information.
  - Utilize AWS Certificate Manager to provision, manage, and renew the SSL/TLS certificates used for encrypting in-transit data.
  - Enforce the use of secure protocols and certificate validation across network connections, including those between internal telco network functions and external interfaces.
- Implement access controls and least privilege:

  - Define granular IAM policies to restrict access to CPNI and PII data based on the principle of least privilege.
  - Leverage AWS Identity and Access Management to create roles and policies that limit access to CPNI and PII data to only the authorized personnel and systems.
  - Implement multi-factor authentication (MFA) for users and systems accessing CPNI and PII data, adding an extra layer of security.
- Enable comprehensive logging and monitoring:
  - Configure AWS CloudTrail to create detailed audit trails of access and usage activities related to CPNI and PII data.
  - Integrate the CloudTrail logs with Amazon CloudWatch and Amazon OpenSearch Service for centralized security monitoring and analysis.
  - Set up Amazon CloudWatch alarms to trigger alerts for suspicious or unauthorized access attempts to CPNI and PII data.
- Establish data governance processes:
  - Develop and enforce data handling policies and procedures aligned with industry regulations and customer data privacy requirements.
  - Integrate the encryption, access control, and logging mechanisms into the organization's overall security framework.
  - Conduct regular audits and assessments to verify the effectiveness of the data protection controls and verify ongoing adherence.

Resources

Key AWS services:

- AWS Key Management Service (KMS)
- AWS Certificate Manager
- AWS Identity and Access Management (IAM)
- AWS CloudTrail
- Amazon CloudWatch
- Amazon OpenSearch Service

Reliability

Reliability in telecommunications networks refers to the ability of the network to provide services without interruption or failure under normal and peak network load conditions. The reliability pillar addresses the ability of a network to recover from infrastructure or service disruptions, dynamically acquire computing and networking resources to meet demand, and mitigate disruptions such as misconfigurations or transient network issues.

Focus areas

- Design principles
- Foundations
- Workload architecture
- Change management
- Failure management

Design principles

- **Scalability:** Telecommunication networks need to dynamically adapt to subscriber growth and usage patterns. The network needs to be elastic and adjusted to the changes. Reliability can be achieved by distributing users and data to multiple resources and scale horizontally instead of vertically. The Communication Services Providers (CSPs) can use micro-services architecture and container-based technologies for agile delivery. This architecture provides high resiliency, better failure handling, efficient use of resources and scale.
- **Failure recovery:** Telecommunication networks need to provide reliable connectivity to the subscribers. Failure can occur from a small component level in software or hardware to a site level. CSPs should have a mechanism to detect service level SLAs which can be translated to technical Key Performance Indicators (KPIs) and try to automatically recover from the failure or have a fail-over to other working resource or site. The KPIs should be used to proactively detect failures and be able to mitigate prior to the occurrence.
- **Application architecture:** The reliability requirements vary from application to application. While some applications are tolerant to latency, other applications such as Ultra Reliable and Low Latency Communication (uRLLC) have stringent latency and jitter requirements. The CSPs prefer to have the network closer to the customer to offer better user experience and to provide

high performance multimedia, real-time applications. The user plane is usually brought closer to the end user, and the control plane network functions can be placed in the central location.

- **Regulatory adherence:** Telecommunication industry is highly regulated and needs to meet regulatory requirements and standards. Special design considerations are needed to meet regulatory requirements. For example, the network should allow emergency calls reliably and to achieve that, resilient systems and networks are needed across the global cloud infrastructure.

Foundations

The foundations section explores the key concepts, definitions, and underlying principles that form the basis of reliability in telecommunications networks. This includes an overview of critical reliability metrics, regulatory requirements, and industry standards that telecom organizations must adhere to verify uninterrupted service delivery.

TELCOREL01: Does your telecommunication network meet local regulatory requirements and service level agreements on the service availability?

Telecom networks operate in a highly regulated environment, where regulatory authorities set guidelines and minimum service availability requirements for CSPs in each country/region. This question focuses on how the telecom organization can meet these regulatory requirements focused on critical service availability.

Best practices

- TELCOREL01-BP01 Conduct regular load and failure tests to validate resilience on node, interface, and infrastructure levels

TELCOREL01-BP01 Conduct regular load and failure tests to validate resilience on node, interface, and infrastructure levels

Implement a comprehensive testing strategy that evaluates network resilience through simulated failures of nodes, interfaces, and infrastructure components. This includes testing in both lower environments and production during maintenance windows to identify potential weaknesses, validate redundancy measures, and verify recovery procedures. The testing should cover both individual component failures and complex failure scenarios that could impact network services.

Desired outcomes

- Identify single points of failure in terms of critical nodes and interfaces that, if they fail, could impact network availability.
- Evaluate network resilience measures and ability to continue providing services to the end user in case of multiple failure scenarios.
- To identify potential performance degradations that may arise due to increased traffic loads.
- Verify effectiveness of redundancy and failover mechanisms.
- Document system behavior and recovery patterns.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

It is essential to regularly test and validate the system's ability to withstand failures, maintaining a reliable and resilient telecom network. This includes defining the test scenarios that cover both network function, interfaces and infrastructure-level failures. Automate the testing and monitoring processes utilizing fault and traffic simulator functions. Define clear success criteria, measurement metrics, and standard documentation practices, to analyze the test results and identify patterns and vulnerabilities. By proactively validating the network's ability to withstand and recover from various failure scenarios, CSPs can verify continuous service availability and meet regulatory requirements for reliability.

Implementation steps

- Define test scenarios for both network function and infrastructure failures:
  - For the network function connectivity failures, Amazon CloudWatch can be used to monitor connectivity, latency, and error rates for your network interfaces and connections.
  - For infrastructure failures, you could leverage Amazon CloudWatch to track the metrics and alarms on your critical AWS resources like EC2 instances, EBS volumes, and network interfaces.
- Create inventory of critical components:
  - Use AWS Config to maintain a comprehensive inventory of your AWS resources and their dependencies.
  - Document interconnections and potential failure impacts using AWS CloudFormation templates or other standard infrastructure as code (IaC) configuration files.
- Design individual component failure tests:

    - Automate failure simulations using AWS Lambda functions or AWS Fault Injection Service to test hardware, software, and network issues.
    - Define expected behavior and recovery procedures using Amazon CloudWatch alarms, and AWS CloudFormation stacks.
- Develop complex failure scenarios:
    - Use auto scaling to model cascading failures, regional outages, and multi-component disruptions.
    - Test resilience by injecting faults across your AWS infrastructure using the AWS Fault Injection Service.
- Establish success criteria:
    - Measure downtime, recovery time, and service quality metrics using Amazon CloudWatch dashboards and alarms.
    - Use Amazon CloudWatch alarms to set thresholds.
- Document and analyze results:
    - Define templates and formats for recording test results, observations, and findings. Include detailed information about test conditions, system behavior, and performance metrics.
    - Review test results to identify common failure modes, systemic issues, and potential vulnerabilities. Look for patterns that might indicate underlying architectural or operational weaknesses.
    - Document and analyze the time required for system recovery under various failure scenarios. Compare results against service level objectives and identify areas for improvement.
    - Assess the effectiveness of automated recovery mechanisms, failover procedures, and list down manual interventions. Identify gaps in current procedures and opportunities for enhancement.
    - Create a prioritized list of areas requiring attention, including system improvements, process changes, and documentation updates.

Resources

Key AWS services:

- Amazon CloudWatch
- AWS Config
- AWS Systems Manager

- AWS Fault Injection Service
- Amazon EventBridge
- AWS CloudFormation

Workload architecture

The workload architecture section focuses on the design patterns and architectural approaches required to build highly available, fault-tolerant, and scalable telecom networks in the cloud. This includes best practices for distributed user plane deployments, control plane redundancy, flexible network function scaling, and cloud-native load balancing solutions.

TELCOREL02: How do you verify high availability and fault tolerance in your network architecture?

Telecom networks underpin essential communication services that individuals and businesses rely on daily. Maintaining the resilience and fault tolerance of these cloud-based networks is paramount as disruptions can hinder vital connections and access to emergency services. Robust network architectures with redundant components, failover mechanisms, and self-healing capabilities are crucial to providing the availability required for modern telecommunication needs.

Best practices

- TELCOREL02-BP01 Deploy user plane functions in a distributed architecture and highly available configuration
- TELCOREL02-BP02 Implement full mesh between control plane and user plane functions
- TELCOREL02-BP03 Implement a flexible network function (NF) design to leverage available infrastructure resources for autoscaling
- TELCOREL02-BP04 Introduce an SCTP load balancer designed for control-plane network functions, carrier-grade performance, and high availability
- TELCOREL02-BP05 Optimize failure recovery timers for the shared tenancy and potential for transient network issues in cloud environments

TELCOREL02-BP01 Deploy user plane functions in a distributed architecture and highly available configuration

Follow a distributed architecture approach to deploy user plane nodes across multiple geographical locations with built-in redundancy. Such a design maintains service continuity through redundant instances and geographical distribution, minimizing the impact of localized failures or outages while optimizing network performance through efficient load distribution.

Desired outcome:

- Achieve high availability on critical network functions.
- Minimized impact from localized failures.
- Optimized network performance.
- Reduced exposure to regional outages.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Implement a distributed and redundant architecture for the user plane functions that can withstand regional service degradations or outages. This involves deploying redundant instances of user plane network functions across multiple geographical locations. Within each site, utilize strategies such as active-active or active-standby configurations, automated health monitoring, and failover strategy. Comprehensive monitoring and data visualization tools provide visibility into the network's performance and availability, enabling rapid identification and response to potential issues. Geographical distribution of these functions, based on user demands and regulatory requirements enhance the resilience and responsiveness of the telecom infrastructure by reducing exposure in case of failure. Automating failover mechanisms, with thresholds and triggers defined by the network management systems, verifies that the network can quickly recover from failures without disrupting service to end users.

Implementation steps

- Deploy redundant instances across multiple Availability Zones:
  - Deploy your 5G network architecture across multiple AWS Regions, Availability Zones, and Outposts where possible.

    - Use AWS Auto Scaling and Amazon CloudWatch to monitor and scale your network function instances based on real-time and projected traffic patterns.
- Implement redundancy patterns:
    - Deploy your user plane functions as Auto Scaling groups with Amazon EC2 instances in an active-active or active-standby configuration.
    - Use Amazon CloudWatch for automated instance health monitoring and failover.
- Establish monitoring:
    - Deploy Amazon CloudWatch to capture and analyze performance and availability metrics across your distributed instances.
- Implement geographical distribution:
    - Choose AWS Regions and Availability Zones based on user distribution, regulatory requirements, and disaster recovery needs.
    - Evaluate AWS network connectivity options, such as AWS Direct Connect and Site-to-Site VPN, to optimize inter-region communication.
- Configure automatic failover:
    - Use Amazon CloudWatch alarms and metrics to define infrastructure and application-level thresholds for initiating automated failover.
    - Integrate with your 5G network management systems to correlate failures and trigger appropriate responses.

Resources

Key AWS services:

- Amazon EKS
- Amazon ECS
- AWS Auto Scaling
- AWS Local Zones

TELCOREL02-BP02 Implement full mesh between control plane and user plane functions

Implement a resilient architecture design where each control plane node can manage user plane functions in a mesh configuration. For each user plane function, one control plane node will be active at a time. The control plane node should be designed to have enough capacity to control the user plane nodes, when needed. This verifies that if one or multiple control plane nodes fail, the remaining nodes can manage user plane functions without service interruption. The design incorporates high-capacity centralized control components with distributed user plane functions.

Desired outcome:

- Enhanced system resilience through redundant connectivity.
- Remove single points of failure in control plane.
- Seamless failover during node failures.
- Maintained service continuity.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

A full mesh connectivity design between the control plane and user plane components is recommended for a highly available Telcom network. Implementing diverse routing paths and continuously monitoring the status of these connectivity routes are key steps. Defining clear failover triggers and thresholds, along with automated recovery procedures, allows the network to quickly respond to and recover from failures. Thorough documentation of these processes, including integration with incident management and troubleshooting guides, further enhances the reliability of the overall system. Comprehensive monitoring and observability solutions provide the necessary visibility into the network's performance and health, enabling proactive identification and mitigation of potential issues.

Implementation steps

- Design mesh topology:
  - Use AWS Transit Gateway to enable connectivity between your control and user plane network functions in a star topology, enabling each control plane node to be able to manage the user plane node, when needed.

    - Verify capacity planning and failure domain considerations using Amazon CloudWatch and AWS Auto Scaling.
- Implement connectivity paths:
    - Deploy control plane and user plane instances across multiple AWS Availability Zones and Regions where possible.
    - Configure diverse routing paths using AWS VPC routing tables and AWS Transit Gateway routing policies.
    - Leverage AWS Direct Connect for physical path separation and high-performance connectivity with the on-premises systems including Access Network (RAN) functions.
- Configure routing policies:
    - Implement routing policies using AWS Transit Gateway route tables and AWS Lambda-based custom routing logic to recover services in case of a control plane node failure.
- Define failover triggers:
    - Use Amazon CloudWatch alarms and metrics to define the conditions and thresholds for triggering automated failover processes.
    - Provide manual override capabilities through AWS Lambda functions or Amazon API Gateway.
- Document procedures:
    - Store the detailed failover and recovery procedures in AWS Systems Manager for secure access and versioning.
    - Integrate the documentation with your incident management processes and provide troubleshooting guides and escalation procedures.

Resources

Key AWS services:

- AWS Transit Gateway
- Amazon VPC
- AWS Lambda
- AWS Direct Connect
- Amazon CloudWatch

TELCOREL02-BP03 Implement a flexible network function (NF) design to leverage available infrastructure resources for autoscaling

Designing a flexible network function design enables the telecom network to dynamically scale its resources based on the availability of suitable instance types in the deployment region. This approach avoids issues like `Insufficient Capacity Error` that can occur when the required instance type is not available, by using alternative instance types that can be provisioned. The architecture should be able to automatically adapt to the most suitable instance types that are present, allowing the network to scale up or down as needed without being constrained by the availability of a specific instance configuration.

Desired outcome:

- Improving the probability of a successful on-demand scale-out in case of traffic surge or failover trigger.
- Network functions can scale seamlessly by utilizing the most suitable instance types that are available in the deployment region.
- Remove the risk of scaling failures due to `Insufficient Capacity Error` when the preferred instance type is not available.
- Verifies continued service availability and responsiveness by maintaining the ability to scale the network functions as needed.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

To build a flexible and scalable telecom network architecture, a modular and containerized approach to network functions is recommended. This involves decomposing network functions into independently scalable components, packaged as container images, and orchestrating them on a Kubernetes-based solution. Dynamic scaling and efficient resource utilization are enabled using a Kubernetes cluster autoscaler, which provisions appropriate instance types based on component requirements. Flexible scaling policies, comprehensive monitoring, failover mechanisms, and thorough documentation verify the network can adapt to changing demands, optimize resource utilization, and maintain overall resilience and reliability.

Implementation steps

- Design a modular and containerized network function architecture:

  - Use a microservices-based approach to decompose network functions into smaller, independently scalable components.
  - Package each network function component as a container image to enable flexibility in deployment and scaling.
- Implement Kubernetes-based orchestration for network functions:
  - Deploy the network function components on Amazon EKS.
  - Integrate the Karpenter open-source Kubernetes cluster autoscaler with your EKS cluster.
  - Configure Karpenter to automatically provision the appropriate instance types based on the resource requirements of the network function components.
- Monitor resource utilization and scaling events:
  - Use Amazon CloudWatch to monitor the performance and resource utilization of the network function components.
  - Analyze scaling events, instance provisioning, and capacity fluctuations to optimize the scaling policies and Karpenter configurations.
- Implement failover and self-healing mechanisms:
  - Configure Kubernetes health checks and readiness probes to detect and replace unhealthy network function instances.
- Document scaling procedures and best practices:
  - Maintain comprehensive documentation on the flexible network function architecture, scaling policies, and Karpenter configuration.
  - Establish guidelines for updating the scaling policies and Karpenter configurations as the instance type availability changes over time.

Resources

Key AWS services:

- Amazon EKS (Elastic Kubernetes Service)
- Karpenter (open-source Kubernetes cluster autoscaler)
- Amazon CloudWatch for monitoring and observability

TELCOREL02-BP04 Introduce an SCTP load balancer designed for control-plane network functions, carrier-grade performance, and high availability

The major public cloud services providers lack native support for load balancing of the Telecom services that rely on control-plane signaling protocols like SCTP (Stream Control Transmission Protocol). It is recommended to introduce a cloud-based SCTP Load Balancer for control plane network functions, designed specifically for telco workloads in the public cloud. Such a solution will provide SCTP protocol awareness, including support for multi-homing and association management, along with carrier-grade performance. The load balancer should also deliver high availability through failover and health checks, maintaining reliable and scalable operation of critical telecommunications components deployed in the public cloud.

Desired outcome:

- Provide a cloud-based SCTP load balancing solution specifically designed for telco control plane workloads.
- Offer built-in support for the SCTP protocol in the network functions, including features like multi-homing, stream multiplexing, and association management.
- Verify high availability through redundancy, failover capabilities, and health monitoring.

Level of risk exposed if this best practice is not established: High

Implementation guidance

To enhance the reliability of telecom networks, an SCTP load balancing solution should be onboarded and deployed. This load balancer should incorporate SCTP-specific features, verify high throughput and low latency, and be implemented with high availability and fault tolerance through redundant deployments and automated failover. Integrate the load balancer with Kubernetes-based network functions, optimize for carrier-grade performance, and implement comprehensive monitoring and automation practices to deliver a reliable and responsive telecom infrastructure.

Implementation steps

- Select an SCTP load balancer:
  - Research and evaluate available SCTP load balancer solutions that offer the required features, such as multi-homing, stream management, and association tracking. Verify the selected

product can scale to handle the throughput and latency requirements of telco control plane workloads.

- Implement high availability and fault tolerance:
  - Deploy the SCTP load balancer across multiple AWS Availability Zones for redundancy.
  - Configure health checks and automated failover mechanisms to verify continuous service availability.
  - Use AWS Auto Scaling to dynamically scale the load balancer instances based on traffic patterns.
- Optimize for carrier-grade performance:
  - Configure the SCTP load balancer to optimize for the specific requirements of telco control plane workloads.
- Implement comprehensive monitoring and observability:
  - Use Amazon CloudWatch to monitor the SCTP load balancer's performance, availability, and error metrics.
  - Configure Amazon CloudWatch alarms and Amazon SNS notifications for proactive alerting and incident management.

Resources

Key AWS services:

- AWS Auto Scaling
- Amazon CloudWatch
- Amazon SNS

TELCOREL02-BP05 Optimize failure recovery timers for the shared tenancy and potential for transient network issues in cloud environments

Many telecom network function vendors implement health monitoring thresholds based on on-premises architectural assumptions, where redundant network interfaces and multiple physical network paths may exist between components. However, when these telecom network functions are deployed in cloud environments, the underlying network is a shared resource with a single connection. This can lead to frequent connection alarms and service disruptions due to transient

network issues or single point of failure events, even though the network function itself may remain functional. It is recommended to proactively engage with telecom ISVs and Network Function developers to optimize the timers and behaviors used for health monitoring and failure recovery. This includes designing network architecture to be resilient to packet losses or higher latencies over the underlying network.

Desired outcome:

- Verify the network functions can effectively detect, react, and recover from network issues and failures.
- Optimize the health monitoring and failure recovery timers to strike the right balance between prompt failure detection and resilience against false positives.
- Use cloud capabilities to supplement vendor-provided failure detection and recovery mechanisms.
- Maintain high availability and service continuity for critical telco network functions despite the dynamic and shared nature of the cloud infrastructure.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Work closely with network function vendors to optimize health monitoring and failure recovery mechanisms for cloud-based deployments. Implement multi-layered monitoring, dynamically adjust failure detection timers, and integrate vendor-provided recovery procedures with cloud-based automation. Validate and test the failure recovery processes, then continuously monitor performance and collaborate with vendors to further refine the configurations over time.

Implementation steps

- Collaborate with network function vendors:
  - Engage with the network function vendors to understand their default health monitoring and failure recovery mechanisms.
  - Work with the vendors to customize and optimize the monitoring and recovery timers based on the cloud solutions and NF failure modes.
- Implement multi-layered health monitoring:
  - Leverage Amazon CloudWatch to set up comprehensive monitoring of the network function instances, including resource utilization, network performance, and application-level metrics.

    - Configure Amazon CloudWatch alarms to detect anomalies and potential failure conditions, with customized thresholds and rules.
    - Implement AWS Lambda-based health check functions to perform more advanced application-level checks, accounting for the specific requirements of the telco network functions.
- Optimize failure detection timers:
    - Work with the network function vendors to fine-tune the health check intervals, failure detection thresholds, and recovery timeouts.
    - Balance the need for prompt failure detection with resilience against transient issues that can trigger false positives.
    - Use Amazon CloudWatch and AWS Lambda to implement dynamic timer adjustments based on observed patterns and cloud infrastructure conditions.
- Enhance failure recovery mechanisms:
    - Integrate the network function vendor-provided recovery procedures with AWS-native capabilities, such as Amazon EC2 Auto Scaling, Amazon EKS, and AWS Lambda.
    - Develop automated recovery workflows using AWS Step Functions or AWS Lambda to handle different types of failures and verify consistent, reliable recovery.
    - Implement rollback and self-healing mechanisms to verify that the network functions can recover to a known good state.
- Validate and test the failure recovery:
    - Use AWS Fault Injection Service to inject various types of failures and network issues into the environment.
    - Monitor the health monitoring and failure recovery mechanisms, and fine-tune the timers and thresholds based on the observed behavior.
    - Document the optimized health monitoring and recovery configurations, including vendor-specific customizations.

Resources

Key AWS services:

- Amazon CloudWatch
- AWS Lambda
- AWS Step Functions

- AWS Auto Scaling
- Amazon EKS
- AWS Fault Injection Service

Change management

The change management section addresses how telecom organizations can manage updates, patches, and other changes to their network and systems in a controlled manner to minimize service disruptions. This includes implementing robust processes for testing, approving, and rolling back changes.

TELCOREL03: How do you manage changes while maintaining the service continuity?

This question addresses how the telecom organization manages changes to their network and systems to maintain uninterrupted service delivery for customers. It involves establishing robust change management processes, testing procedures, and rollback capabilities to minimize service disruptions during system scaling, updates, and upgrades.

Best practices

- TELCOREL03-BP01 Use a controlled change management mechanism for software updates in production
- TELCOREL03-BP02 Design a stateful, session-aware load balancing solution for cloud network functions (CNFs) to enable scaling and resilience

TELCOREL03-BP01 Use a controlled change management mechanism for software updates in production

Follow a controlled change management approach that blocks automatic updates in production environments while maintaining changes are thoroughly tested in lower environments. This includes establishing proper testing procedures, implementing structured change management, and maintaining comprehensive rollback capabilities to minimize service disruption during system updates.

Desired outcome:

- Minimized risk of service disruptions from updates.
- Validated changes before production deployment.
- Controlled deployment processes.
- Documented change procedures.
- Verified update effectiveness.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Prevent automatic software updates in production environments on the infrastructure and network function levels. Implement a controlled change management process for the system and service updates. Identify update mechanisms, configure update policies to disable automatic updates while maintaining security adherence, and document detailed manual update procedures. Establish comprehensive change control processes, including approval workflows, testing requirements, and documentation standards. Maintain a centralized tracking system for updates and develop robust rollback capabilities with automated and manual recovery procedures. Continuously monitor the update process, define clear criteria for triggering rollbacks, and validate the effectiveness of the recovery mechanisms.

Implementation steps

- Disable automatic updates in production:
  - Use AWS Systems Manager Patch Manager to manage and control operating system and application updates across your AWS environment.
  - Use AWS Config to monitor and audit the configuration of your systems, including the update settings.
- Configure update policies:
  - Use AWS Config to define and enforce specific configurations for disabling automatic updates while maintaining security.
  - Use AWS Systems Manager Parameter Store to centrally manage and store your updated policies and configurations.
- Establish change controls:
  - Use AWS Config Rules to implement comprehensive change management processes, including approval workflows, testing requirements, and documentation standards.

  - Use AWS CloudTrail to track and audit update-related activities across your AWS environment.
- Create a tracking system:
  - Use AWS Config to track and monitor update-related changes, including pending updates, approved changes, and implementation status.
  - Use Amazon DynamoDB to store update-related data and provide an audit trail of activities.
- Maintain rollback procedures:
  - Use AWS Systems Manager Automation to design and implement the rollback mechanisms, including both automated and manual rollback capabilities.
  - Use Amazon CloudWatch to establish the clear criteria for determining when rollback procedures should be initiated, including both automated triggers and manual decision points.
  - Use AWS Config to verify the effectiveness of the rollback procedures by validating system stability and functionality after rollback completion.

Resources

Key AWS services:

- AWS Systems Manager
- AWS Config
- Amazon CloudWatch

TELCOREL03-BP02 Design a stateful, session-aware load balancing solution for cloud network functions (CNFs) to enable scaling and resilience

Cloud network functions (CNFs) such as AMF, SMF, and UPF in 5G architectures are typically designed to be horizontally scalable. However, these components maintain subscriber session state, and scaling events (up and down) can disrupt session continuity, leading to call drops, session re-establishment delays, or degraded user experience. To address this challenge, it is recommended to design a stateful, session-aware load balancing solution for CNFs that enables seamless scaling and resilience. This can be achieved by collaborating with the CNF vendors and relevant CNF communities to explore how cloud-based services can be adapted to support seamless session handling.

Desired outcome:

- Provide a highly available and scalable load balancing solution that can maintain session state for telco network functions deployed as cloud-based workloads.
- Verify seamless failover and state preservation during scaling events or instance failures, minimizing service disruptions for end users.
- Enable efficient load distribution and resource utilization by intelligently routing traffic based on current capacity and workload patterns.
- Integrate with Kubernetes-based deployment models and service abstractions used by telco CNFs.
- Deliver carrier-grade performance and availability to meet the stringent requirements of telco control plane and signaling workloads.

Implementation guidance

Design a stateful load balancing architecture that maintains session state and seamless failover, leveraging in-memory data stores or distributed databases to replicate session data across instances. Integrate the load balancing solution with Kubernetes-based cloud-based network functions, providing a standardized interface and allowing CNFs to configure load balancing parameters. Implement a highly available and fault-tolerant design, with automated failover and dynamic scaling capabilities. Enhance observability and monitoring to maintain visibility into the load balancing system's performance, errors, and configuration changes. Automate the deployment and lifecycle management of the load balancing solution to verify consistency.

Implementation steps

- Design a stateful load balancing architecture:
  - Use AWS application load balancing services, such as Application Load Balancer (ALB) and Network Load Balancer (NLB), to handle the traffic routing and load distribution.
  - Implement session stickiness and connection tracking mechanisms to maintain session state and verify seamless failover.
  - Utilize Amazon ElastiCache (For example, Redis) or Amazon DynamoDB to store and replicate session data across the load balancing instances.
- Integrate with Kubernetes-based CNFs:
  - Develop a Kubernetes Ingress Controller or Gateway API implementation to provide a standardized interface for telco CNFs to leverage the stateful load balancing solution.

  - Implement custom annotations or custom resource definitions (CRDs) to allow telco CNFs to configure load balancing parameters, such as session timeouts and stickiness.
  - Verify the load balancing solution can discover and register telco CNF instances through Kubernetes service discovery mechanisms.
- Optimize for carrier-grade performance:
  - Configure the load balancing instances to leverage high-performance AWS instance types, such as C5 or M5 families, to handle the throughput and latency requirements of telco workloads.
  - Use AWS Nitro System-based instances for optimized networking and CPU performance.
  - Integrate with AWS Direct Connect and AWS Global Accelerator to provide low-latency connections and Regional load distribution.
- Implement high availability and fault tolerance:
  - Deploy the load balancing solution across multiple AWS Availability Zones for redundancy and resilience.
  - Configure health checks, automated failover, and connection draining to verify seamless failover during instance or Availability Zone failures.
  - Leverage AWS Auto Scaling to dynamically scale the load balancing capacity based on traffic patterns and resource utilization.
- Enhance observability and monitoring:
  - Integrate the load balancing solution with Amazon CloudWatch for comprehensive monitoring of performance metrics, error rates, and health status.
  - Implement Amazon CloudWatch dashboards and alarms to provide visibility into load balancing behavior and trigger alerts for anomalies or potential issues.
  - Use AWS CloudTrail and AWS Config to track configuration changes and maintain an audit trail of load balancing activities.
- Automate deployment and lifecycle management:
  - Package the load balancing solution as an AWS CloudFormation template or Terraform configuration for consistent and repeatable deployment.
  - Integrate the solution with AWS CodePipeline and AWS CodeBuild for automated CI/CD workflows, including testing and deployment.
  - Use AWS Config and AWS Systems Manager to manage the configuration of the load balancing components and verify adherence with defined policies.

Resources

Key AWS services:

- Application Load Balancer (ALB)
- Network Load Balancer (NLB)
- Amazon ElastiCache (For example, Redis)
- Amazon DynamoDB
- Kubernetes Ingress Controller
- AWS Auto Scaling
- Amazon CloudWatch
- AWS CloudTrail
- AWS Config
- AWS CodePipeline
- AWS CodeBuild
- AWS Systems Manager

Failure management

The failure management section focuses on proactive measures to block and rapidly respond to various failure scenarios in the telecom network. This includes implementing signaling firewalls to protect against DDoS attacks, deploying redundant control and user plane architectures, and establishing default configurations to bypass billing systems in the event of outages.

TELCOREL04: How do you block and manage failures in your telco network?

Maintaining reliable service in a complex, mission-critical telecom network is paramount. Telecom operators must proactively block and rapidly respond to various failure scenarios, from individual component outages to large-scale disruptive events.

Best practices

- TELCOREL04-BP01 Implement the Bypass Mode on peripheral network devices such as firewalls and network probes while connecting to the cloud

- TELCOREL04-BP02 Implement dual network planes for signaling or control plane
- TELCOREL04-BP03 Implement default configuration to bypass billing and charging services in case the system is down

TELCOREL04-BP01 Implement the Bypass Mode on peripheral network devices such as firewalls and network probes while connecting to the cloud

Implementing bypass mode on network peripheral devices like firewalls and network probes is crucial for maintaining service continuity during device failures or power outages. This mechanism automatically creates a direct connection between incoming and outgoing network ports when a device fails, maintaining that critical network traffic continues to flow without interruption. By deploying bypass-capable devices while connecting to the cloud-based network infrastructure, telecom operators can block service disruptions that could occur during device failures, maintenance windows, or unexpected outages. This approach is particularly important for maintaining high availability in telecommunications networks where continuous service is essential for customer satisfaction and regulatory adherence.

Desired outcome:

- Verify continuous network traffic flow during device level failures.
- Minimize service disruptions during maintenance or unexpected outages.
- Maintain high availability for critical network services.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Design your network architecture to include redundant paths and automated routing updates that can quickly respond to device failures. Implement comprehensive health monitoring and automated recovery procedures to minimize service disruptions. Establish clear operational procedures for managing bypass events, including notification workflows and post-event analysis to block recurring issues.

Regular testing of the bypass functionality is crucial to verify it will work as expected during actual failures. This includes conducting controlled failure scenarios during maintenance windows

and validating that traffic continues to flow without interruption. Document bypass events and maintain historical data to identify patterns and potential improvements to the bypass implementation.

Implementation steps

- Deploy AWS Network Firewall with fail-open configurations across multiple Availability Zones to verify continuous traffic flow during failures.
- Configure AWS Transit Gateway for centralized network management with automatic routing failover capabilities.
- Use Amazon CloudWatch to monitor device health and performance metrics, with automated alerts for potential failures on AWS services involved.

Resources

Key AWS services:

AWS Transit Gateway

Amazon CloudWatch

TELCOREL04-BP02 Implement dual network planes for signaling or control plane

A resilient network architecture implementing two independent and redundant signaling networks that operate in parallel. Each signaling plane handles messages and controls network resources independently, with logical and physical separation to verify that failures in one plane do not impact the other plane's operations.

Desired outcome:

- Enhanced fault tolerance through dual planes.
- Independent operation capability.
- Seamless failover between planes.
- Maintained service continuity.
- Isolated failure domains.

- Verified redundancy effectiveness.
- Documented recovery procedures.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Design a resilient network architecture with separate signaling planes for control plane traffic, maintaining comprehensive logical and physical separation. Enforce strict network access controls and security zoning to maintain the separation. Configure robust failover mechanisms, with automated triggers, recovery procedures, and comprehensive health monitoring across the planes. Test the failover functionality under various failure scenarios and document the recovery steps. Deploy advanced monitoring and observability tools to track the performance, availability, and security of both signaling planes, establish key metrics and alerting thresholds, and regularly review the monitoring strategy for continuous improvement.

Implementation steps

- Design separate signaling planes:
  - Use AWS Transit Gateway and AWS Network Firewall to implement a vendor-neutral strategy for logical and physical separation between control and user plane traffic. Configure AWS Network ACLs and security groups to enforce network zoning, traffic classification, and security controls.
- Verify physical separation:
  - Use AWS Config to specify and enforce the technical requirements for network isolation, including network segments, addressing, and security controls. Use AWS Config Rules to monitor and enforce requirements affecting the separation.
  - Deploy the separate network infrastructure for each plane using Amazon VPC and AWS Direct Connect. Utilize AWS Transit Gateway and AWS Network Firewall to implement network virtualization and segmentation.
  - Configure AWS Network Firewall and Amazon VPC security groups to enforce strict network access controls and security policies. Define and enforce the security zones and trust boundaries using AWS IAM and AWS Service Control Policies.
- Configure failover mechanisms:
  - Use Amazon CloudWatch alarms and AWS Lambda functions to establish the conditions and thresholds for automated failover.

- Deploy Amazon CloudWatch and AWS Lambda to monitor the health of the planes and trigger the appropriate failover responses.
- Use AWS Fault Injection Service to regularly test the failover functionality under various failure conditions. Document the test procedures and success criteria in AWS Systems Manager.
- Create detailed recovery procedures for different failure scenarios. Include troubleshooting guides and contact information.

- Monitor plane status:
  - Use Amazon CloudWatch to monitor the performance, availability, and security metrics for both planes.
  - Establish key performance indicators using Amazon CloudWatch that reflect service health, including latency, throughput, error rates, and resource utilization. Set appropriate thresholds and baselines.
  - Configure Amazon CloudWatch alarms and Amazon SNS to implement a notification system for critical events and threshold violations. Define alert severity levels and response procedures.
  - Regularly evaluate monitoring strategy effectiveness and adjust based on operational experience. Update metrics and thresholds as needed.

Resources

Key AWS services:

- Amazon VPC
- AWS Transit Gateway
- Amazon CloudWatch
- AWS Config
- AWS Lambda
- AWS Systems Manager

TELCOREL04-BP03 Implement default configuration to bypass billing and charging services in case the system is down

Implementing a default configuration to bypass billing and charging services in a telecom network maintains service continuity for customers in the event of a failure or outage in the billing or

charging system. The network is configured with a failover mechanism that continues providing critical communication services while bypassing the charging/billing component if they become unavailable or unresponsive. This bypass mode allows customers to continue accessing and using the telecom services without interruption, even when the billing or charging systems are down.

Desired outcome:

- Verify customer experience and service availability take precedence over billing and charging functions.
- Allow customers to continue using services while their usage and session data is temporarily stored or cached within the network.
- Process cached usage data once the billing and charging systems are restored, verifying revenue is not lost.

Implementation guidance

Implement a default configuration on the control plane functions that allows the network to bypass the billing and charging systems in the event of a failure or outage. Configure the network functions to continue call setup or session establishment even when the charging system is unresponsive. When the billing or charging systems are restored, they process the cached usage data to verify revenue is not lost despite the temporary bypass of these systems.

Implementation steps

- Define bypass triggers and thresholds:
  - Use monitoring tools to detect failures or unresponsiveness in the billing and charging systems
  - Establish clear criteria and thresholds for triggering the bypass mode, such as connection timeouts, error rates, or system availability metrics
- Implement bypass configuration:
  - Configure network functions with a default method to continue call setup or session setup in case the charging system is unresponsive.
- Use data records and call records for offline charging/billing:
  - Once the billing or charging systems are back online, the cached usage data can be processed and the end user is billed accordingly, verifying that revenue is not lost despite the temporary bypass of these systems.

Resources

Key AWS services:

- Amazon CloudWatch for monitoring and alerting
- AWS Lambda for implementing custom bypass logic and data processing

Performance efficiency

The performance efficiency pillar includes the ability to use computing resources efficiently to meet system requirements, and to maintain that efficiency as demand changes and technologies evolve.

The performance efficiency pillar provides an overview of design principles, best practices, and questions. You can find prescriptive guidance on implementation in the Performance Efficiency pillar whitepaper.

Focus areas

- Design principles
- Architecture selection
- Compute and hardware
- Networking and content delivery
- Process and culture
- Data management

Design principles

1. **Automate processes:** Leverage automation to remove manual tasks and reduce the likelihood of human error. Use tools to automate infrastructure provisioning, monitoring, and incident response workflows.
2. **Monitor and log system performance:** Implement comprehensive monitoring and logging of telecom workloads. Regularly analyze logs to identify patterns, trends, and areas for improvement.
3. **Establish performance baselines:** Define and establish performance baselines for telecom workloads, allowing for the identification of anomalies and deviations from expected behavior. Set baselines and monitor them continuously.
4. **Iterate and improve:** Continuously evaluate and improve telecom workloads by implementing feedback loops and incorporating learnings from incidents and operational data. Track changes and manage configurations.
5. **Implement proactive incident response:** Develop and maintain playbooks for incident response and recovery and use automation to minimize the impact of failures.

6. **Apply AI/ML for operational insights:** Utilize AI and ML technologies to analyze operational data and gain insights that can optimize network performance, capacity planning, and customer support.
7. **Prioritize documentation and knowledge sharing:** Maintain up-to-date documentation of telecom workloads, including architectural diagrams, runbooks, and operational procedures.
8. **Foster a culture of operational excellence:** Encourage a culture of continuous improvement, where team members are empowered to identify and address operational challenges, and learn from failures. Perform regular reviews and measure the alignment with best practices.

Network design principles

When designing a high-performance, resilient, and secure network on AWS that connects to telcos, several key capabilities and services should be considered. Here are some of the essential components:

1. Establish dedicated network connections between your on-premises data center or telecom network and your cloud provider. Use a reliable, low-latency, and high-bandwidth connection that bypasses the public internet.
2. Implement a highly available and scalable Domain Name System (DNS) service with advanced routing features like latency-based routing, Geo DNS, and health checks. Direct traffic to the most appropriate resources based on location and availability.
3. Implement load balancing to distribute incoming traffic across multiple targets, such as EC2 instances or containers, to verify high availability and fault tolerance.
4. AWS Wavelength: Deploy 5G applications at the edge of the mobile network using AWS Wavelength. Wavelength Zones are AWS infrastructure deployments embedded within the telco's data centers, providing ultra-low latency connectivity to mobile users and devices.
5. AWS Local Zones: Deploy 5G, Cable, and Wireline networks using AWS Local Zones.

Architecture selection

TELCOPERF01: How do you determine which geographic regions to host your telco infrastru cture?

The strategic placement of telecom infrastructure across geographic regions is crucial for optimizing network performance, maintaining regulatory adherence, and meeting customer needs. This consideration involves evaluating factors such as latency requirements, data sovereignty laws, disaster recovery planning, and proximity to end-users. The choice of hosting locations can significantly impact service quality, operational efficiency, and the overall resilience of the telecom network.

Best practices

- TELCOPERF01-BP01 Use Edge locations to deploy latency sensitive telco network workloads such as RAN and user-plane nodes
- TELCOPERF01-BP02 Select the infrastructure regions to deploy telco workloads based on performance requirements and regulatory considerations
- TELCOPERF01-BP03 Deploy the control plane network functions on centralized locations to meet high scalability and agility requirements

TELCOPERF01-BP01 Use Edge locations to deploy latency sensitive telco network workloads such as RAN and user-plane nodes

Deploying latency-sensitive telco network workloads like Radio Access Network (RAN) and user plane nodes in edge locations is recommended. Placing these components closer to the end users at the edge of the network minimizes latency and improves performance for real-time, latency-critical applications and services.

Desired outcome:

- Place latency-sensitive telco workloads like Radio Access Network (RAN) and user plane nodes closer to end users at the network edge.
- Minimize latency and improve performance for real-time, latency-critical telco applications and services.
- Enhance the overall user experience by providing faster response times.

Common anti-patterns:

- Deploying telco workloads in centralized data centers without considering proximity to end users.
- Failing to use edge computing capabilities to offload latency-sensitive tasks.

- Neglecting to optimize network infrastructure for low-latency requirements.

Benefits of establishing this best practice:

- Reduced latency for critical telco services like voice, video, and real-time data processing.
- Improved quality of experience (QoE) for end users by providing faster response times.
- Ability to scale edge resources independently to handle localized traffic spikes.
- Enhanced resilience through distributed architecture and reduced single points of failure.
- Reduced bandwidth consumption and costs by processing data closer to the source.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Deploying latency-sensitive telco network workloads like Radio Access Network (RAN) and user-plane nodes in edge locations is a crucial strategy for optimizing network performance and improving user experience. By placing these components closer to the end users at the edge of the network, telco operators can minimize latency and verify that critical services like voice, video, and real-time data processing are delivered with low latency and high responsiveness.

Edge computing solutions like AWS Wavelength and AWS Local Zones provide the infrastructure to host these latency-sensitive telco workloads close to the end users. By leveraging these services, telco operators can offload compute-intensive tasks to the edge, reducing the need to backhaul traffic to centralized data centers and maintaining that end users experience the best possible service.

When implementing this best practice, telco operators should carefully assess the specific latency requirements of their services and the geographic distribution of their customer base. This will inform the selection of appropriate edge locations and the workloads that should be deployed at the edge versus in centralized data centers.

Implementation steps

- Identify the telco workloads and services that have the most stringent latency requirements, such as RAN and user-plane functions.
- Analyze the geographic distribution of your customer base and the locations where these latency-sensitive services are consumed.

- Evaluate the availability of AWS Wavelength Zones and AWS Local Zones that can host these workloads closer to the end users.
- Use Amazon EC2 instances in the selected AWS Wavelength Zones or AWS Local Zones to deploy the identified latency-sensitive telco workloads, verifying they are optimized for low-latency performance.
- Configure AWS Auto Scaling to automatically scale the EC2 instances hosting the latency-sensitive workloads based on traffic patterns and resource utilization.
- Use Amazon CloudWatch to monitor the performance of the edge-deployed workloads and trigger auto scaling or relocation of resources as needed to maintain optimal user experience.

Resources

Key AWS services:

- AWS Wavelength
- AWS Local Zones

TELCOPERF01-BP02 Select the infrastructure regions to deploy telco workloads based on performance requirements and regulatory considerations

Selecting the appropriate infrastructure regions for deploying telco workloads is crucial for optimizing network performance and maintaining regulatory adherence. When choosing regions, consider factors such as latency requirements, data sovereignty laws, and the geographical distribution of your customer base. For latency-sensitive applications like voice services or real-time video, prioritize regions closer to your end users. Also, consider regional regulations regarding data storage and processing, especially for customer information and call records. By strategically selecting infrastructure regions, you can enhance service quality, reduce latency, and maintain adherence with local regulations, improving the overall user experience and operational efficiency of your telco services.

Desired outcome:

- Select infrastructure regions that minimize latency for end users by placing workloads closer to customer base.

- Verify adherence with local data sovereignty and regulatory requirements for data storage and processing.
- Improve overall service quality and user experience by optimizing network performance and adhering to regional regulations.

Common anti-patterns:

- Deploying telco workloads in a single infrastructure region without considering latency or regulatory factors.
- Failing to evaluate the geographic distribution of customers and placing workloads far from the majority of users.
- Disregarding data sovereignty laws and regulatory requirements when selecting infrastructure Regions.

Benefits of establishing this best practice:

- Reduced latency and improved performance for latency-sensitive telco services like voice and video.
- Adherence with local data privacy and sovereignty regulations, avoiding fines and legal issues.
- Ability to better scale infrastructure and resources to meet demand in specific geographic regions.
- Enhanced user experience and customer satisfaction through optimized network performance.
- Improved operational efficiency by aligning infrastructure placement with business and regulatory needs.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Selecting the appropriate infrastructure regions for deploying telco workloads is crucial for optimizing network performance and maintaining regulatory adherence. When choosing regions, telco operators should consider several key factors:

- **Latency requirements:** For latency-sensitive applications like real-time voice and video services, prioritize infrastructure regions that are geographically closer to the end-user base to minimize

latency. This may involve deploying certain workloads at the network edge using services like AWS Wavelength or AWS Local Zones.

- **Data sovereignty and regulatory adherence:** Evaluate regional regulations regarding data storage, processing, and retention, especially for customer information and call records. Deploy workloads in regions that allow you to maintain adherence with local data sovereignty laws.
- **Geographic distribution of customers:** Analyze the geographic distribution of your customer base and align your infrastructure regions accordingly. This verifies that most users can access telco services with optimal performance.

By strategically selecting infrastructure regions that balance latency, regulatory adherence, and geographic coverage, telco operators can enhance service quality, reduce latency, and maintain adherence to local regulations, improving the overall user experience and operational efficiency of their services.

Implementation steps

- Use the AWS Region Selector tool to assess the latency, data sovereignty laws, and regulatory requirements for different AWS Regions that align with your telco customer base.
- Create AWS VPCs in the selected Regions that meet your performance needs, configuring subnets, routing tables, and security groups accordingly.
- Deploy telco workloads in the appropriate AWS Regions, using services like Amazon EC2, Amazon EKS, and AWS Fargate for the latency-sensitive components closer to end users and the regulatory-sensitive components in compatible Regions.
- Configure AWS Direct Connect or Site-to-Site VPN connections to establish dedicated network links between your on-premises telco infrastructure and the selected AWS Regions.
- Use Amazon CloudWatch and AWS CloudTrail to continuously monitor the network performance and security of your telco workloads across the AWS infrastructure.

Resources

Key AWS services:

- AWS Wavelength
- AWS Local Zones

TELCOPERF01-BP03 Deploy the control plane network functions on centralized locations to meet high scalability and agility requirements

Running the control plane nodes in centralized geographical locations is a strategy used to meet the demands for high scalability and mobility. Centralizing the control plane components in strategically placed data centers enables efficient resource management, load balancing, and fault tolerance across the distributed system. This approach allows the system to better handle sudden traffic increases, support many client nodes, and provide reliable failover mechanisms in case of individual node failures.

Desired outcome:

- Deploy the control plane components of the telco network in centralized, strategically placed data centers.
- Achieve high scalability and mobility for the control plane to handle sudden increases in traffic and support a large number of client nodes.
- Provide reliable failover mechanisms and fault tolerance for the control plane functions.

Common anti-patterns:

- Distributing control plane components across multiple geographic locations without a centralized strategy.
- Failing to scale control plane resources appropriately to meet demand spikes.
- Lacking robust failover and redundancy mechanisms for critical control plane functions.

Benefits of establishing this best practice:

- Improved scalability and agility to handle dynamic traffic patterns and sudden surges in demand.
- Enhanced fault tolerance and reliability through centralized control plane architecture.
- Efficient resource management and load balancing across the distributed telco network.
- Simplified operations and maintenance of the control plane components.
- Reduced operational costs through optimized resource utilization.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Running the control plane nodes of a telco network in centralized geographical locations is a strategic approach to meet the demands for high scalability and mobility. By consolidating the control plane components in strategically placed data centers, telco operators can achieve efficient resource management, load balancing, and fault tolerance across the distributed system. This centralized control plane architecture allows the network to better handle sudden traffic increases, support many client nodes, and provide reliable failover mechanisms in case of individual node failures.

The control plane components, responsible for functions like authentication, authorization, and mobility management, can be scaled more effectively in a centralized deployment, maintaining that the overall network can adapt to changing demands and maintain consistent service quality. Moreover, the centralized control plane design simplifies operations and maintenance, as the critical functions are managed in a consolidated manner. This approach enables telco operators to leverage advanced monitoring, automation, and orchestration capabilities to optimize the performance and reliability of the control plane.

When implementing this best practice, telco operators should carefully select the geographical locations for the control plane data centers, considering factors such as network latency, data sovereignty, and disaster recovery planning. The control plane components should be deployed with high availability and redundancy mechanisms to verify continuous operation and seamless failover in case of infrastructure failures.

Implementation steps

- Identify the key control plane components and functions within your telco network architecture.
- Deploy the control plane components in centralized AWS Regions, using Amazon EC2 instances or Amazon EKS clusters to host the highly available and redundant control plane infrastructure.
- Configure AWS Route 53 for DNS and service discovery, enabling efficient communication between the control plane and user plane components.
- Implement Amazon CloudWatch and AWS CloudTrail to monitor the health, performance, and security of the centralized control plane deployment.
- Set up AWS Lambda functions to automate the scaling, failover, and recovery processes for the control plane components, maintaining efficient resource utilization and rapid response to changes in demand.
- Regularly test the control plane failover and disaster recovery procedures using AWS services like Amazon EC2 Auto Scaling, AWS CloudFormation, and AWS Backup.

Resources

Key AWS services:

- Amazon EC2
- Amazon EKS
- Amazon Route 53
- AWS CloudFormation

Compute and hardware

TELCOPERF02: How do you select the appropriate compute solution for your telco workloads ?

The selection of compute resources must account for specific telecommunications needs such as deterministic performance, ultra-low latency, hardware acceleration capabilities, and support for specialized telecom protocols. This includes consideration of bare metal, virtualized, and containerized environments, as well as specific instance types optimized for telecom workloads like Network Function Virtualization (NFV) and Radio Access Network (RAN) processing.

Best practices

- TELCOPERF02-BP01 Implement containers to achieve optimal performance and resource utilization

TELCOPERF02-BP01 Implement containers to achieve optimal performance and resource utilization

Implementing containers in your telco architecture is a crucial step towards achieving optimal performance and resource utilization. When adopting this approach, it is essential to carefully consider your instance type and hardware selection. For example, User Plane Function (UPF) components require specific performance capabilities to handle high data plane throughput, while Radio Access Network (RAN) elements may benefit from specialized hardware accelerators.

Desired outcome:

- Achieve greater agility, scalability, and efficiency in the telco network infrastructure using containers.
- Optimize resource utilization and manage the compute footprint dynamically to handle traffic spikes and evolving demands.
- Use advanced container orchestration features to improve the reliability and fault tolerance of telco workloads.

Common anti-patterns:

- Relying solely on traditional virtual machine-based deployments without adopting containers.
- Failing to consider the specific performance and resource requirements of telco workloads when implementing containers.
- Lacking the appropriate container orchestration solution and features needed to manage the scale and complexity of telco networks.

Benefits of establishing this best practice:

- Improved performance and throughput for latency-sensitive telco workloads like User Plane Function (UPF) and Radio Access Network (RAN).
- Better resource utilization and cost optimization through right-sizing and dynamic scaling of container-based telco components.
- Increased agility and flexibility to deploy, update, and manage telco network functions.
- Enhanced reliability and fault tolerance through container orchestration capabilities.
- Simplified operations and reduced maintenance overhead for the telco infrastructure.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Implementing containers in your telco architecture is a crucial step towards achieving optimal performance and resource utilization. When adopting this approach, it is essential to carefully consider your instance type and hardware selection to meet the specific requirements of telco workloads.

For example, User Plane Function (UPF) components require specific performance capabilities to handle high data plane throughput, while Radio Access Network (RAN) elements may benefit from

specialized hardware accelerators. Optimizing your compute footprint is another key aspect of container implementation, involving careful management of resource allocation to avoid over-provisioning and verify efficient utilization.

Selecting the optimal container control plane and orchestration characteristics is also crucial for telco workloads. This is particularly important for control plane components, where factors such as Transactions Per Second (TPS) and session requirements play a significant role. Choose a container orchestration solution that can handle the scale and complexity of telco networks, providing features like service mesh for improved communication between microservices, and robust monitoring and logging capabilities for effective troubleshooting and performance optimization. By carefully implementing containers with these considerations in mind, telco operators can achieve greater agility, scalability, and efficiency in their network infrastructure, leading to improved service quality and reduced operational costs.

Implementation steps

- Use Amazon EKS to deploy your telco workloads in a managed Kubernetes environment, allowing you to take advantage of container-based architectures.
- Select EC2 instance types that are optimized for the performance and resource requirements of your telco user plane and control plane components, such as high-memory or compute-optimized instances.
- Integrate AWS Fargate, a serverless compute engine for containers, to manage the scaling and provisioning of the container infrastructure without the need to manage underlying EC2 instances.
- Configure AWS App Mesh, a service mesh for microservices, to enhance communication, observability, and security between the containerized telco components.
- Use Amazon CloudWatch and AWS X-Ray to monitor the performance, resource utilization, and health of your containerized telco workloads, triggering auto scaling actions as needed.

Resources

Key AWS services:

- Amazon EKS (Elastic Kubernetes Service)
- AWS Fargate
- Amazon EC2 Spot Instances

- AWS App Mesh

Networking and content delivery

TELCOPERF03: How do you select the appropriate network infrastructure for your telco workloads?

Network infrastructure is the backbone of telecom operations to verify optimal performance and reliability. This involves evaluating options for both physical and virtual networking components, including switches, routers, load balancers, and software-defined networking (SDN) solutions. The selection process must also account for future growth, emerging technologies like 5G and edge computing, and the ability to support diverse services ranging from voice and data to IoT and multimedia applications.

Best practices

- TELCOPERF03-BP01 Implement dedicated network infrastructure for control and user plane functions
- TELCOPERF03-BP02 Deploy hardware acceleration solutions for enhanced packet processing and network performance
- TELCOPERF03-BP03 Implement network timing synchronization and distribution mechanisms to verify service consistency

TELCOPERF03-BP01 Implement dedicated network infrastructure for control and user plane functions

Separating network infrastructure for control and user plane functions (CUPS) is essential for optimizing telco workload performance. Control plane infrastructure should be designed with high availability and security features to handle signaling traffic effectively, while user plane infrastructure must be optimized for high throughput and low latency to manage subscriber data. This separation enables independent scaling of each plane based on specific requirements while maintaining secure and efficient inter-plane communication, leading to better resource utilization and improved service quality.

Desired outcome:

- Separate the network infrastructure for control plane and user plane functions to enhance performance and scalability.
- Verify the control plane infrastructure is designed for high availability and security to handle signaling traffic effectively.
- Optimize the user plane infrastructure for high throughput and low latency to manage subscriber data efficiently.

Common anti-patterns:

- Deploying control plane and user plane functions on a shared network infrastructure without separation.
- Failing to consider the distinct performance and scaling requirements of the Control and User Plane components.
- Neglecting to design the appropriate high availability and security mechanisms for the control plane network.

Benefits of establishing this best practice:

- Improved overall network performance and service quality by optimizing the control and user plane components independently.
- Enhanced scalability and the ability to scale each plane based on specific requirements.
- Increased security and reliability of the control plane functions through dedicated network infrastructure.
- Better resource utilization and cost optimization by aligning the network infrastructure with the distinct needs of each plane.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Separating the network infrastructure for control plane and user plane functions (CUPS) is a critical best practice for optimizing the performance and scalability of telco workloads. This architectural approach recognizes the distinct requirements and characteristics of these two fundamental components of telco networks.

The control plane infrastructure should be designed with high availability and security features to handle the signaling traffic and control-related functions effectively. This includes maintaining robust failover mechanisms, efficient load balancing, and advanced security measures to protect the critical control plane elements.

In contrast, the user plane infrastructure must be optimized for high throughput and low latency to manage the subscriber data and user traffic efficiently. This may involve the use of specialized hardware acceleration solutions, such as SmartNICs and FPGAs, to offload intensive packet processing tasks and improve overall network performance.

By implementing this separation of the control and user plane networks, telco operators can scale each plane independently based on their specific requirements, leading to better resource utilization and improved service quality. This approach also enables more efficient inter-plane communication, as the dedicated network infrastructure can be tailored to the needs of each component.

When deploying this best practice, telco operators should consider the overall network architecture, the expected traffic patterns, and the specific performance and scaling requirements of the control and user plane functions. This may involve the use of services like AWS Wavelength or AWS Local Zones to strategically place the user plane components closer to the end users for reduced latency.

Implementation steps

- Create separate VPCs within your AWS environment to host the control plane and user plane components, configuring appropriate subnets, routing tables, and security groups for each.
- Establish secure communication between the control plane and user plane VPCs using AWS PrivateLink, AWS Transit Gateway, or Site-to-Site VPN connections.
- Configure AWS Direct Connect or AWS Outposts to provide dedicated, high-bandwidth network connectivity for the user plane infrastructure, optimizing for low-latency and high-throughput.
- Use AWS Network Load Balancer to distribute traffic across the user plane components, and AWS Application Load Balancer to manage the control signaling traffic for the Control Plane.
- Implement Amazon CloudWatch and AWS CloudTrail to monitor the performance, security, and health of the separated control plane and user plane network infrastructure.

Resources

Key AWS services:

- Amazon VPC
- Amazon EC2
- AWS Wavelength
- AWS Local Zones
- AWS Security Groups
- AWS Direct Connect

TELCOPERF03-BP02 Deploy hardware acceleration solutions for enhanced packet processing and network performance

Implementing hardware acceleration solutions such as SmartNICs and FPGAs is crucial for meeting the demanding performance requirements of modern telco workloads. These solutions offload intensive packet processing tasks from the main CPU, significantly improving network performance and reducing latency. By integrating hardware acceleration with virtualized network functions (VNFs) and containerized network functions (CNFs), operators can achieve the performance levels required for 5G networks while maintaining the flexibility of cloud-based architectures.

Desired outcome:

- Improve the network performance of telco workloads by offloading intensive packet processing tasks from the main CPU.
- Achieve lower latency and higher throughput for telco services that require high-performance networking.
- Enhance the overall efficiency and scalability of the telco network infrastructure by using hardware acceleration technologies.

Common anti-patterns:

- Relying solely on software-based packet processing without considering hardware acceleration solutions.
- Failing to integrate hardware acceleration with virtualized or containerized network functions.
- Neglecting to optimize the configuration and tuning of hardware acceleration technologies for telco-specific requirements.

Benefits of establishing this best practice:

- Significant performance improvements for latency-sensitive telco services like 5G networks.
- Increased throughput and reduced CPU utilization for network-intensive workloads.
- Better scalability and the ability to handle higher traffic volumes without compromising performance.
- Improved energy efficiency and reduced operational costs by offloading networking tasks to specialized hardware.
- Enhanced reliability and fault tolerance using dedicated networking acceleration components.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Implementing hardware acceleration solutions, such as SmartNICs and FPGAs, is a crucial strategy for meeting the demanding performance requirements of modern telco workloads. These technologies offload intensive packet processing tasks from the main CPU, significantly improving network performance and reducing latency.

By integrating hardware acceleration with virtualized network functions (VNFs) and containerized network functions (CNFs), telco operators can achieve the performance levels required for 5G networks while maintaining the flexibility and agility of cloud-based architectures. This approach allows telco workloads to fully leverage the capabilities of the underlying hardware, maintaining that critical services like real-time communications, video streaming, and edge computing applications can deliver exceptional service to end users.

When deploying hardware acceleration solutions, telco operators should carefully evaluate the specific requirements of their network functions and the available hardware options. This may involve testing and benchmarking different acceleration technologies to determine the optimal fit for their telco workloads. Additionally, it is important to verify that the hardware acceleration

Integration of hardware acceleration with virtualized and containerized network functions is another key aspect of this best practice. Telco operators should work closely with their technology partners and solution providers to verify seamless integration and optimization of the hardware acceleration capabilities within their telco network architecture.

Implementation steps

- Evaluate the availability of Amazon EC2 instances with FPGA accelerators, such as the F1 instance family, to offload network packet processing tasks.

- Integrate the hardware-accelerated EC2 instances with your virtualized or containerized telco network functions, leveraging the AWS Nitro System for optimized performance.
- Configure the hardware acceleration components to optimize their performance for your specific telco protocols, traffic patterns, and data processing requirements.
- Use Amazon CloudWatch and AWS CloudTrail to monitor the utilization and performance of the hardware acceleration solutions, adjusting as needed to maintain optimal network efficiency.
- Consider deploying your latency-sensitive telco workloads in AWS Wavelength Zones or AWS Local Zones to take advantage of the proximity to end users and potential hardware acceleration capabilities

Resources

Key AWS services:

- Amazon EC2 Instances with FPGA
- AWS Nitro System
- AWS Wavelength
- AWS Local Zones

TELCOPERF03-BP03 Implement network timing synchronization and distribution mechanisms to verify service consistency

Establishing robust network timing synchronization across distributed infrastructure is critical for maintaining consistent service quality in telco workloads. Proper timing mechanisms verify smooth handovers, accurate billing, and reliable service delivery, particularly in edge computing scenarios. Deploy timing solutions that comply with 3GPP specifications and support future network evolution, while considering geographic distribution requirements and the need for precise synchronization between network elements at different locations.

Desired outcome:

- Establish robust network timing synchronization across the distributed telco infrastructure.
- Verify smooth handovers, accurate billing, and reliable service delivery, particularly in edge computing scenarios.

- Deploy timing solutions that comply with 3GPP specifications and support future network evolution.

Common anti-patterns:

- Failing to implement dedicated network timing synchronization mechanisms.
- Deploying timing solutions that do not meet the stringent requirements of telco workloads.
- Neglecting to consider the geographic distribution of network elements when designing the timing architecture.

Benefits of establishing this best practice:

- Accurate billing and charging processes by maintaining precise synchronization of network events.
- Enhanced reliability and seamless handovers between network elements, even in edge computing deployments.
- Adherence with industry standards and regulations, enabling future network evolution and interoperability.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Establishing robust network timing synchronization across the distributed telco infrastructure is critical for maintaining consistent service quality and reliability. Proper timing mechanisms verify smooth handovers, accurate billing, and reliable service delivery, particularly in edge computing scenarios where network elements may be geographically dispersed.

telco operators should deploy timing solutions that comply with 3GPP specifications and support future network evolution, such as the transition to 5G. These solutions should be able to provide precise synchronization between network elements at different locations, considering the geographic distribution of the infrastructure and the need for accurate timestamping of network events.

By implementing a comprehensive network timing synchronization and distribution strategy, telco operators can verify that critical services like voice, video, and real-time data processing are delivered with consistent quality, and that billing and charging processes are accurate and reliable.

This level of timing precision is essential for maintaining a seamless user experience and complying with industry standards and regulations.

When designing the network timing architecture, telco operators should consider factors such as the geographic spread of their infrastructure, the specific requirements of their telco services, and the compatibility with existing and emerging network protocols and standards. This may involve the use of dedicated timing distribution mechanisms, such as IEEE 1588 Precision Time Protocol (PTP) or Global Navigation Satellite System (GNSS) technologies, to verify accurate and reliable time synchronization across the network**.**

Implementation steps

- Use AWS Outposts to deploy on-premises network equipment that supports precision timing protocols like IEEE 1588 Precision Time Protocol (PTP) and Global Navigation Satellite System (GNSS).
- Configure AWS Wavelength Zones to provide the necessary timing synchronization and distribution capabilities for your telco workloads deployed at the network edge.
- Leverage AWS Ground Station, a fully managed satellite ground station service, to access GNSS timing data and distribute it to your telco network components across different Regions.
- Implement Amazon CloudWatch and AWS CloudTrail to monitor the performance and reliability of the network timing synchronization, triggering alerts and automated responses for deviations from the expected behavior.

Resources

Key AWS services:

- AWS Outposts
- AWS Wavelength
- AWS Ground Station

Process and culture

TELCOPERF04: How do you select the appropriate monitoring solution for your telco workloads?

Effective monitoring solutions for telecom workloads require comprehensive visibility into both infrastructure and service performance metrics to maintain and meet service level agreements (SLAs).

Best practices

- TELCOPERF04-BP01 Implement AI and ML-powered monitoring solutions to gain insights into network health and performance
- TELCOPERF04-BP02 Monitor deviations from nominal performance parameters using system alerts and automated responses

TELCOPERF04-BP01 Implement AI and ML-powered monitoring solutions to gain insights into network health and performance

Implementing advanced monitoring solutions that use artificial intelligence (AI) and machine learning (ML) technologies is crucial for optimizing telco network performance and operational efficiency. By using AI solutions, telecom operators can gain deep insights into network behavior, identify patterns, and predict potential issues before they impact service quality. These AI-driven monitoring solutions can analyze vast amounts of network data in real-time, enabling more accurate capacity planning, proactive network optimization, and improved customer support. This approach not only enhances overall network performance but also reduces operational costs and improving customer satisfaction by minimizing service disruptions and enabling faster problem resolution.

Desired outcome:

- Use AI/ML technologies to gain deep insights into telco network behavior and performance.
- Enable more accurate capacity planning, proactive network optimization, and improved customer support through advanced data analysis.
- Enhance overall network performance and reduce operational costs by minimizing service disruptions and enabling faster problem resolution.

Common anti-patterns:

- Relying solely on traditional, rule-based monitoring solutions without leveraging AI/ML capabilities.

- Failing to integrate AI/ML-powered monitoring with the broader telco network management and automation workflows.
- Neglecting to continuously train and improve the AI/ML models to adapt to evolving network conditions and new use cases.

Benefits of establishing this best practice:

- Improved network health visibility and the ability to predict and block performance issues.
- Faster identification and resolution of network problems through automated root cause analysis.
- Enhanced customer experience by proactively addressing service quality and reliability concerns.
- Optimized resource utilization and capacity planning based on data-driven insights.
- Reduced operational costs through improved efficiency and reduced service disruptions.

Risk Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Implementing advanced monitoring solutions that use AI/ML technologies is crucial for optimizing telco network performance and operational efficiency. By using AI/ML solutions, telco operators can gain deep insights into network behavior, identify patterns, and predict potential issues before they impact service quality.

These AI-driven monitoring solutions can analyze vast amounts of network data in real-time, enabling more accurate capacity planning, proactive network optimization, and improved customer support. This approach not only enhances overall network performance but also reduces operational costs and improving customer satisfaction by minimizing service disruptions and enabling faster problem resolution.

When deploying AI/ML-powered monitoring solutions, telco operators should consider integrating them with their broader network management and automation workflows. This allows the insights generated by the monitoring system to be seamlessly leveraged for tasks like dynamic resource scaling, automated incident response, and predictive maintenance.

It is also important to establish processes for continuously training and refining the AI/ML models used in the monitoring solution. As the telco network evolves, the monitoring capabilities must adapt to new use cases, changing traffic patterns, and emerging technologies to maintain their effectiveness.

Implementation steps

- Use Amazon SageMaker AI to build, train, and deploy custom machine learning models that can analyze telco network data and provide insights into performance, anomalies, and predictive maintenance.
- Integrate Amazon Kinesis Data Streams and Amazon Data Firehose to ingest and pre-process the telco network data in real-time, feeding it into the SageMaker AI-powered AI/ML models.
- Configure Amazon CloudWatch to trigger automated actions, such as scaling resources or initiating incident response workflows, based on the insights generated by the AI/ML models.
- Use the AWS IoT Core service to collect and analyze telemetry data from edge devices and network elements, feeding this information into the AI/ML-powered monitoring solution.
- Continuously review and refine the AI/ML models using the latest telco network data and feedback from network operations, verifying the monitoring solution remains effective over time.

Resources

Key AWS services:

- Amazon SageMaker AI
- Amazon Kinesis
- Amazon CloudWatch
- AWS IoT Core

TELCOPERF04-BP02 Monitor deviations from nominal performance parameters using system alerts and automated responses

Effective monitoring of telecom workloads requires establishing and tracking performance baselines to quickly identify anomalies and deviations from expected behavior. By leveraging observability solutions, operators can implement comprehensive monitoring and alerting systems that continuously track system performance against established baselines. This approach enables automated responses to specific events and early detection of potential issues, assisting to maintain optimal network performance and reducing the risk of service disruptions. Setting up these monitoring systems with appropriate thresholds and alert mechanisms verifies that operations teams can proactively address performance issues before they impact service quality.

Desired outcome:

- Establish comprehensive monitoring and alerting systems to continuously track telco workload performance against defined baselines.
- Quickly identify anomalies and deviations from expected behavior to enable proactive issue resolution.
- Implement automated responses and remediation actions to address specific performance-related events and maintain optimal network operations.

Common anti-patterns:

- Lacking well-defined performance baselines and thresholds for telco workloads.
- Relying solely on manual monitoring and incident response without leveraging automated systems.
- Failing to integrate monitoring and alerting capabilities with the broader telco network management and operations workflows.

Benefits of establishing this best practice:

- Early detection and blocking performance issues that could impact service quality and customer experience.
- Reduced mean time to identify and resolve network problems, minimizing service disruptions.
- Improved operational efficiency and reduced manual intervention through automated responses to specific events.
- Better alignment between network performance and business objectives by continuously monitoring against established baselines.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Effective monitoring of telco workloads requires establishing and tracking performance baselines to quickly identify anomalies and deviations from expected behavior. By leveraging observability solutions, telco operators can implement comprehensive monitoring and alerting systems that continuously track system performance against the established baselines.

This approach enables automated responses to specific events and early detection of potential issues, assisting to maintain optimal network performance and reducing the risk of service disruptions. Setting up these monitoring systems with appropriate thresholds and alert mechanisms verifies that operations teams can proactively address performance issues before they impact service quality.

When implementing this best practice, telco operators should consider integrating the monitoring and alerting capabilities with their broader network management and automation workflows. This allows the insights and actions generated by the monitoring system to be seamlessly leveraged for tasks like dynamic resource scaling, automated incident response, and predictive maintenance.

Additionally, telco operators should establish processes for regularly reviewing and updating the performance baselines and alert thresholds. As the network evolves, traffic patterns change, and modern technologies are introduced, the monitoring system must adapt to maintain its effectiveness in identifying and addressing performance-related issues.

Implementation steps

- Define the key performance metrics and baselines for your telco workloads in Amazon CloudWatch, considering the requirements of different services and network components.
- Configure Amazon CloudWatch alarms to track deviations from the established performance baselines, setting appropriate thresholds and triggering mechanisms.
- Integrate the CloudWatch alarms with AWS Lambda functions to automate the response and remediation actions for specific performance-related events, such as scaling resources or rerouting traffic.
- Use AWS CloudTrail to maintain a comprehensive audit trail of configuration changes and actions taken by the automated monitoring and response system.
- Regularly review the performance baselines and alert thresholds in Amazon CloudWatch, adjusting as the telco network and workloads evolve.

Resources

Key AWS services:

- Amazon CloudWatch
- AWS CloudTrail
- AWS Lambda

- Amazon SNS

Data management

TELCOPERF05: How do you select the appropriate storage solution for your telco workloads?

Selecting the appropriate storage solution for telco workloads is crucial for managing performance and cost requirements. The key is to categorize your telco data based on access patterns, regulatory needs, and business value, and then match the storage solution accordingly.

Best practices

- TELCOPERF05-BP01 Implement data lifecycle management strategies for IVR and call logs to optimize storage costs and performance

TELCOPERF05-BP01 Implement data lifecycle management strategies for IVR and call logs to optimize storage costs and performance

Implementing effective data lifecycle management for Interactive Voice Response (IVR) systems and call logs is crucial in telecommunications environments. This strategy involves categorizing data based on access patterns, regulatory requirements, and business value, then moving it through different storage tiers to balance performance and cost-effectiveness. By automating the transition of data from high-performance storage for frequently accessed records to more cost-effective archival solutions for older data, telecommunications providers can maintain optimal system performance while managing storage costs efficiently.

Desired outcome:

- Categorize telco data (for example, IVR recordings and call logs) based on access patterns, regulatory requirements, and business value.
- Implement automated data lifecycle management to transition data between different storage tiers, balancing performance, and cost-effectiveness.
- Maintain optimal system performance for frequently accessed data while managing long-term storage costs efficiently.

Common anti-patterns:

- Storing telco data on high-performance storage without considering lifecycle and access patterns.
- Failing to implement data tiering and archiving strategies to manage the growing volume of data.
- Neglecting to align data retention policies with regulatory requirements and business needs.

Benefits of establishing this best practice:

- Reduced storage costs by moving less frequently accessed data to more economical storage solutions.
- Improved system performance and responsiveness for applications that rely on real-time access to telco data.
- Adherence with industry regulations and internal data retention policies through automated data lifecycle management.
- Enhanced operational efficiency by automating the movement and management of telco data across storage tiers.
- Better utilization of storage resources through rightsizing and optimizing the storage footprint.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Implementing effective data lifecycle management for Interactive Voice Response (IVR) systems and call logs is crucial in telecommunications environments. This strategy involves categorizing data based on access patterns, regulatory requirements, and business value, then moving it through different storage tiers to balance performance and cost-effectiveness.

By automating the transition of data from high-performance storage for frequently accessed records to more cost-effective archival solutions for older data, telecommunications providers can maintain optimal system performance while managing storage costs efficiently. This approach verifies that mission-critical data, such as recent call logs and IVR recordings, are readily available in high-performance storage, while historical data that is less frequently accessed is stored in a more cost-effective manner.

When implementing this best practice, telco operators should consider the specific retention requirements for different types of telco data, as defined by industry regulations and internal policies. The data lifecycle management strategy should be designed to automatically move data between storage tiers based on these retention policies, minimizing the risk of data loss or unauthorized access.

Additionally, telco operators should integrate the data lifecycle management processes with their broader data management and analytics workflows. This allows the insights gained from analyzing telco data to inform the optimization of the storage strategy and the categorization of data into appropriate tiers.

Implementation steps

- Use Amazon S3 to store the different types of telco data (for example, IVR recordings, call logs and billing records), categorizing them into appropriate S3 buckets based on access patterns, regulatory requirements, and business value.
- Configure S3 Lifecycle policies to automatically transition data between S3 storage classes, such as moving older data from S3 Standard to Amazon Glacier or S3 Glacier Deep Archive, based on the defined data retention policies.
- Integrate Amazon EFS and Amazon EBS to provide high-performance file storage and block storage for the frequently accessed telco data, complementing the long-term archival capabilities of S3.
- Use AWS Storage Gateway to seamlessly connect your on-premises telco infrastructure with the appropriate AWS storage services, enabling a hybrid storage architecture that optimizes for performance and cost.
- Implement Amazon CloudWatch and AWS CloudTrail to monitor the data lifecycle management processes, triggering alerts and automated actions for deviations from the defined policies.

Resources

Key AWS services:

- Amazon S3 (including Amazon Glacier and S3 Glacier Deep Archive)
- Amazon EFS
- Amazon EBS
- AWS Storage Gateway

Cost optimization

The cost optimization pillar includes the continual process of refinement and improvement of a system over its entire lifecycle. From the initial design of your first proof of concept to the ongoing operation of production workloads, adopting the practices in this paper will allow you to build and operate cost-aware systems that achieve business outcomes and minimize costs, thus allowing your business to maximize its return on investment.

Focus areas

- Design principles
- Expenditure and usage awareness
- Cost-effective resources
- Data transfer
- Data storage
- Interoperability
- Manage demand and supply resources

Design principles

In addition to the general design principles from the cost optimization pillar of the Well-Architected Framework whitepaper, the following design principles can assist you to optimize the costs of your telco network infrastructure.

- **Measure cost at every layer of the network infrastructure**: The costs to develop, operate, and maintain the network have significant impact on the financial viability of a telco. Therefore, it is important to understand and track the infrastructure costs at each layer - core, distribution, access, edge - so you can identify areas that may require cost optimization.

Expenditure and usage awareness

TELCOCOST01: How do you want to measure the cost of your telco workload across various domains?

This question focuses on understanding and measuring the cost of your telco workload across various domains. This is an important aspect of cost optimization, as you need visibility into your different cost drivers to identify opportunities for savings.

TELCOCOST01-BP01 Implement attribution of cost to each telco domain hosted on the AWS (access, core, edge, OSS, and BSS)

Telcos with workloads running on AWS can take advantage of AWS' detailed cost and usage reporting to attribute costs to each of their domains (like access, core, edge, OSS, and BSS). By tagging resources appropriately, costs can be allocated to each domain.

Desired outcome:

- Gain visibility into the cost breakdown of your telco workload across different domains.
- Identify high-cost areas and make informed decisions to optimize spending.
- Improve cost transparency and accountability across your telco organization.

Common anti-patterns:

- Lack of cost tagging and tracking for resources, leading to unclear cost attribution.
- Siloed cost management, with different teams responsible for distinct domains.
- Reliance on high-level consolidated billing reports without granular cost breakdowns.

Benefits of establishing this best practice:

- Enables data-driven cost optimization decisions.
- Improves cross-team collaboration and accountability for cost management.
- Facilitates chargeback or show back models for internal cost allocation.
- Supports regulatory adherence and auditing requirements.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Accurately attributing costs to different telco domains is essential for effective cost optimization. By using AWS's detailed cost and usage reporting capabilities, you can gain visibility into the

specific costs associated with each of your telco domains. This information can assist you to identify areas of high spending and make informed decisions to optimize costs.

Implementation steps

- Verify you have enabled detailed billing and cost allocation reports in your AWS account.
- Create AWS Cost and Usage tags for your resources to categorize costs by domain.
- Apply the appropriate tags to your AWS resources, such as EC2 instances, S3 buckets, and Lambda functions.
- Configure AWS Cost Explorer to analyze and visualize your costs by the tagged domains.
- Set up AWS Budgets and alerts to monitor and receive notifications on cost trends for each domain.
- Regularly review the cost allocation reports and optimize costs in high-spend domains.

Resources

Key AWS services:

- AWS Cost and Usage Reports
- AWS Cost Explorer
- AWS Budgets

Cost-effective resources

TELCOCOST02: How are you choosing the right compute solution for your telco Network Functions?

This question focuses on selecting the most cost-optimized compute solutions for your telco Network Functions.

Best practices

- TELCOCOST02-BP01 Implement dynamic CNF sizing and scaling strategies based on actual subscriber demands and usage patterns

- TELCOCOST02-BP02 Choose the most efficient compute resource for your Network Function

TELCOCOST02-BP01 Implement dynamic CNF sizing and scaling strategies based on actual subscriber demands and usage patterns

In traditional on-premises environments, CNF infrastructure is typically sized for peak capacity expected over the hardware lifetime. To optimize costs in cloud environments, organizations should implement dynamic sizing strategies that align infrastructure capacity with actual subscriber demands. This includes utilizing cloud-based auto scaling capabilities, implementing rightsizing based on usage patterns, and designing CNF architectures that support horizontal scaling. The approach should balance performance requirements with cost efficiency while maintaining the ability to handle growth in subscriber demand through elastic infrastructure scaling.

Desired outcome:

- Optimize the cost of your cloud-based network functions (CNFs) by right-sizing resources based on actual usage.
- Verify CNF performance meets service-level objectives while minimizing over-provisioning.
- Achieve cost savings by dynamically scaling CNF resources up and down in response to changing demand.

Common anti-patterns:

- Static, one-size-fits-all provisioning of CNF resources without considering variable demand.
- Over-provisioning of CNF resources to handle peak capacity, leading to excess costs during off-peak periods.
- Lack of visibility into CNF resource utilization and performance, hindering effective scaling decisions.

Benefits of establishing this best practice:

- Significant cost savings by aligning CNF resources with actual subscriber needs.
- Improved CNF performance and reliability by adapting to changing demand patterns.
- Enhanced operational efficiency through automated scaling and resource management.

- Increased agility in responding to growth or seasonal fluctuations in subscriber base.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implementing dynamic sizing and scaling strategies for your cloud-based Network Functions (CNFs) is crucial for optimizing costs in a cloud-based telco environment. Unlike traditional on-premises deployments, the cloud provides the opportunity to right-size your CNF resources based on actual subscriber demands and usage patterns, rather than provisioning peak capacity.

By using cloud-based auto scaling capabilities, you can dynamically adjust the resources allocated to your CNFs, verifying performance requirements are met while minimizing over-provisioning and associated costs. This approach allows you to scale resources up and down in response to changes in subscriber demand, improving cost-efficiency without compromising the user experience.

Implementation steps

- Profile the performance and resource utilization characteristics of your cloud-based network functions (CNFs).
- Use AWS Auto Scaling to automatically scale CNF resources up and down based on real-time metrics like CPU, memory, and network utilization.
- Configure Amazon CloudWatch alarms to trigger auto scaling actions when thresholds are breached.
- Use Amazon EC2 Auto Scaling groups with a dynamic target tracking scaling policy to maintain optimal performance at the lowest cost.
- Analyze historical usage patterns and seasonality to set appropriate scaling thresholds and policies.
- Continuously monitor and refine your autoscaling configurations to verify they adapt to changing traffic patterns.

Resources

Key AWS services:

- AWS Auto Scaling
- Amazon CloudWatch

- AWS Lambda

TELCOCOST02-BP02 Choose the most efficient compute resource for your Network Function

When implementing cloud-based network functions (CNFs), it is important to benchmark your applications on different compute types to determine the optimal balance of performance and cost-efficiency. CNFs have diverse compute requirements, so gravitating to the lowest cost option may result in poor performance. Older compute architectures and hardware can hamper efficiency and drive-up expenses over time.

Desired outcome:

- Identify the most cost-effective compute resources that meet the performance requirements of your cloud-based network functions (CNFs).
- Optimize the balance between cost and performance for your CNF workloads.
- Avoid over-provisioning or under-provisioning of compute resources, which can lead to increased costs.

Common anti-patterns:

- Selecting compute resources solely based on the lowest cost, without considering performance requirements.
- Relying on outdated compute architectures and hardware that are less efficient and cost-effective.
- Lack of benchmarking and profiling of CNF workloads on different compute options.

Benefits of establishing this best practice:

- Significant cost savings by selecting the most optimal compute resources for your CNFs.
- Improved CNF performance and reliability by matching compute capabilities to workload needs.
- Enhanced operational efficiency by right-sizing compute resources to avoid over-provisioning.
- Increased agility in responding to changing compute requirements for your CNF workloads.

Level of risk exposed if this best practice is not established: High

Implementation guidance

When implementing cloud-based network functions (CNFs), it is crucial to carefully evaluate the compute resource options to strike the right balance between cost and performance. CNFs often have diverse compute requirements, and simply selecting the lowest-cost compute option may result in poor performance and sub-optimal outcomes.

By benchmarking your CNF applications on different compute types, you can identify the most efficient resources that meet your performance needs. This may involve evaluating various AWS compute services, such as Amazon EC2 instances with different processor architectures, memory configurations, and storage options. Older compute architectures and hardware can also hamper efficiency and drive-up expenses over time, so it is important to consider the long-term costs and benefits of your compute choices.

The goal is to optimize the cost-performance tradeoff for your CNF workloads, verifying you are not over-provisioning or under-provisioning compute resources, which can lead to increased costs.

Implementation steps

- Profile the performance and resource utilization characteristics of your cloud-based network functions (CNFs).
- Benchmark your CNF applications on a variety of AWS compute instances, including different processor architectures, memory configurations, and storage options.
- Analyze the performance and cost data to identify the most efficient compute resources that meet your CNF's requirements.
- Consider the long-term implications of your compute choices, evaluating factors like energy efficiency, hardware lifecycle, and future performance trends.
- Implement mechanisms to dynamically adjust the compute resources allocated to your CNFs based on changing demands and workload characteristics.
- Continuously monitor and optimize your CNF compute resource allocations to verify cost-effectiveness while maintaining performance.

Resources

Key AWS services:

- Amazon EC2

- AWS Graviton Processor
- AWS Compute Optimizer

Data transfer

TELCOCOST03: How are you optimizing data transfer costs between and within your telco network infrastructure?

This question focuses on optimizing data transfer costs within and between the various components of a telco's network infrastructure. Data transfer costs can be a significant driver of overall telco expenditure, so it is important to carefully manage and optimize these costs. The best practices discussed address strategies like implementing strategies to intelligently route traffic and leveraging edge zones and microservices to minimize data transfers across the network.

Best practices

- TELCOCOST03-BP01 Use edge-zones and services to implement cloud workloads made up of containers and microservices

TELCOCOST03-BP01 Use edge-zones and services to implement cloud workloads made up of containers and microservices

Telecom companies can take advantage of the cloud's ability to replicate resources across Availability Zones (AZs) within the same Region to reduce data transfer costs. When launching infrastructure, telecoms should distribute workloads across multiple Availability Zones but use transit gateways to interconnect them. This allows for high availability without paying for data transfer between Availability Zones, since this is free.

Desired outcome:

- Reduce data transfer costs by minimizing traffic between regional and edge locations.
- Improve application performance and user experience by serving content from the network edge.
- Achieve high availability and fault tolerance through distributed, microservices-based architectures.

Common anti-patterns:

- Monolithic application architectures with centralized data and processing.
- Lack of edge computing capabilities to bring services closer to users.
- Inefficient communication patterns between distributed components, leading to high data transfer costs.

Benefits of establishing this best practice:

- Significant reduction in data transfer costs by utilizing edge locations and private connectivity.
- Enhanced application performance and responsiveness through edge-based processing and content delivery.
- Improved resilience and fault tolerance through a distributed, microservices-based design.
- Increased agility in scaling and deploying new features and services.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Adopting a cloud-based, microservices-based architecture and using edge computing capabilities can be highly effective in optimizing data transfer costs for your telco workloads. By distributing your application components across multiple Availability Zones and edge locations, you can reduce the need for data to traverse long distances across the network, thereby minimizing data transfer charges.

AWS services like Amazon ECS, Amazon EKS, AWS PrivateLink, and AWS Global Accelerator can assist you to implement this approach. By using these services, you can deploy your containerized microservices in a distributed manner, use private connectivity between them to avoid public data transfers, and route user traffic to the closest edge location, further reducing data transfer costs.

Implementation steps

- Design your application architecture using a microservices approach with containers.
- Deploy your containerized microservices across multiple Availability Zones within an AWS Region using Amazon ECS or Amazon EKS.
- Use AWS PrivateLink to enable private connectivity between your microservices without incurring data transfer costs.

- Use AWS Global Accelerator to route user traffic to the closest edge location, minimizing latency and data transfer costs.
- Implement service discovery using AWS Cloud Map to enable efficient communication between microservices.
- Monitor your application's traffic patterns and adjust your edge and regional deployment strategy to optimize data transfer costs.

Resources

Key AWS services:

- Amazon ECS
- Amazon EKS
- AWS PrivateLink
- AWS Global Accelerator
- AWS Cloud Map

Data storage

TELCOCOST04: How are you optimizing the data storage costs for your network infrastructure?

This question focuses on optimizing the data storage costs associated with a telco's network infrastructure. Data storage can be a significant cost driver for telcos, as they need to manage large volumes of data from various network elements, customer records, billing systems, and more.

Best practices

- TELCOCOST04-BP01 Choose the appropriate type of storage for network functions backups, metrics, KPIs and the event records to reduce costs
- TELCOCOST04-BP02 Use ETSI ENI based architectures to implement intelligent network slicing

TELCOCOST04-BP01 Choose the appropriate type of storage for network functions backups, metrics, KPIs and the event records to reduce costs

Telecom companies generate and store massive amounts of data to support their business and customer operations. However, data is accessed or needed at various frequencies. By aligning storage choices with data access needs and lifecycles, Telecoms can significantly reduce their storage costs. For actively accessed data that needs high performance, flash storage or cache are good options despite their higher costs. For medium-term data that is accessed occasionally, lower-cost options like object storage, SAN or NAS storage are suitable. For long-term archive data with infrequent access, cold storage options like tape or cloud archival storage are the most cost-efficient.

Desired outcome:

- Align storage choices with the access patterns and retention requirements of different data types.
- Achieve cost savings by utilizing the most cost-effective storage options for each data category.
- Verify appropriate performance and durability characteristics for the various data workloads.

Common anti-patterns:

- One-size-fits-all storage approach, with the same storage solution used for each data type.
- Overreliance on high-cost storage options for data that does not require frequent access.
- Lack of visibility and control over storage usage and costs across the organization.

Benefits of establishing this best practice:

- Significant cost savings by optimizing storage costs based on data access patterns.
- Improved storage efficiency and resource utilization.
- Enhanced data management through appropriate retention and tiering strategies.
- Increased agility in responding to changing storage requirements.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Telecom companies generate and store large volumes of data to support their business and customer operations. However, data is accessed or needed at different frequencies. By aligning storage choices with data access needs and lifecycles, telecoms can significantly reduce their storage costs.

For actively accessed data that needs high performance, flash storage or cache are good options despite their higher costs. For medium-term data that is accessed occasionally, lower-cost options like object storage, SAN, or NAS storage are suitable. For long-term archive data with infrequent access, cold storage options like tape or cloud archival storage are the most cost-efficient.

Implementation steps

- Categorize your data based on access frequency, performance requirements, and retention needs (for example, active, medium-term, and long-term archive).
- For active data that requires high performance, use Amazon EBS Provisioned IOPS SSD volumes or Amazon EFS.
- For medium-term data that is accessed occasionally, use Amazon S3 Intelligent-Tiering or Amazon EFS.
- For long-term archive data with infrequent access, use Amazon Glacier or Amazon Glacier Deep Archive.
- Implement lifecycle policies to automatically transition data between storage tiers as access patterns change.
- Monitor storage usage and costs and adjust your storage tiering strategy as needed to optimize costs.

Resources

Key AWS services:

- Amazon EBS
- Amazon EFS
- Amazon S3
- Amazon Glacier
- Amazon Glacier Deep Archive

TELCOCOST04-BP02 Use ETSI ENI based architectures to implement intelligent network slicing

The European Telecommunications Standards Institute (ETSI) Experiential Networked Intelligence (ENI) is an architectural framework that defines standards for cognitive network management and implementation of 5G use cases based on environmental context and user requirements. It allows Telcos to take advantage of cloud-based technologies like network slicing, service mesh, and microservices to build more agile and automated networks.

Desired outcome:

- Improve network resource utilization and efficiency through dynamic, context-aware network slicing.
- Enhance the customer experience by automatically allocating network resources based on user and application requirements.
- Achieve cost savings by right-sizing network capacity to match evolving demands.

Common anti-patterns:

- Static, one-size-fits-all network provisioning without considering variable user and application needs.
- Lack of real-time visibility into network conditions and user/application demands.
- Inability to rapidly adapt network resource allocation in response to changing requirements.

Benefits of establishing this best practice:

- Optimized network resource utilization and efficiency.
- Improved customer experience through tailored network capabilities.
- Reduced network operating costs by aligning capacity with actual demands.
- Increased network agility and responsiveness to evolving requirements.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

The European Telecommunications Standards Institute (ETSI) Experiential Networked Intelligence (ENI) architecture provides a framework for implementing intelligent, context-aware network slicing. By using the key components of the ENI architecture, such as the policy management function and the context awareness function, telecoms can build cloud-based, microservices-based network slicing solutions that dynamically allocate resources based on user and application needs.

This approach enables telecoms to improve network resource utilization and cost-effectiveness by right-sizing network capacity to match actual demands, rather than provisioning for peak requirements. Additionally, the context-aware nature of the ENI architecture allows the network to automatically adapt to changing conditions, user behavior, and application requirements, enhancing the overall customer experience.

Implementation steps

- Familiarize yourself with the ETSI ENI architecture and its key components, such as the Policy Management Function (PMF) and the Context Awareness Function (CAF).
- Design your network slicing architecture using the ETSI ENI principles, including the use of microservices, service mesh, and cloud-based technologies.
- Use AWS services like Amazon EKS, AWS App Mesh, and AWS Lambda to implement the ENI-based network slicing components.
- Develop policies and rules within the PMF to dynamically allocate network resources based on user and application requirements.
- Use the CAF to gather contextual information about network conditions, user behavior, and application demands to drive intelligent resource allocation.
- Continuously monitor the performance and cost-effectiveness of your ENI-based network slicing implementation and adjust as needed.

Resources

Key AWS services:

- Amazon EKS
- AWS App Mesh
- AWS Lambda

Interoperability

TELCOCOST05: How are you handling the interoperability of your radio access network?

This question focuses on addressing the interoperability of a telco's radio access network (RAN) in a cost-effective manner. Interoperability of the RAN is an important consideration for cost optimization, as it can impact the overall flexibility, scalability, and vendor dependence of the network infrastructure.

Best practices

- TELCOCOST05-BP01 Explore open interface-based technology like RAN or vRAN to reduce network-related costs

TELCOCOST05-BP01 Explore open interface-based technology like RAN or vRAN to reduce network-related costs

To examine O-RAN in more detail and understand the best practices for O-RAN architectures on AWS, see Open Radio Access Network Architecture on AWS.

Desired outcome:

- Achieve cost savings through the adoption of open interface-based RAN or vRAN technologies.
- Increase flexibility and vendor independence in the radio access network.
- Improve scalability and agility in deploying new RAN capabilities.

Common anti-patterns:

- Reliance on proprietary, vendor-locked RAN solutions that limit flexibility and drive-up costs.
- Lack of consideration for open interface-based RAN architectures and their potential benefits.
- Inability to rapidly adapt the RAN to changing business and technology requirements.

Benefits of establishing this best practice:

- Significant cost savings through the adoption of open interface-based RAN/vRAN technologies.

- Increased flexibility and vendor independence, reducing lock-in and enabling faster innovation.
- Improved scalability and agility in deploying new RAN capabilities to meet evolving demands.
- Enhanced ability to use cloud-based technologies and services to manage the RAN.

Level of risk exposed if this best practice is not established: Medium

Implementation guidance

Open Radio Access Network (O-RAN) is an industry initiative that defines open interface specifications for the radio access network (RAN). By adopting O-RAN and vRAN architectures, telecoms can reduce network-related costs and increase flexibility in their RAN infrastructure.

The O-RAN approach allows telecoms to decouple the different RAN components, such as the Radio Unit (O-RU), Distributed Unit (O-DU), and Central Unit (O-CU), and use diverse vendors and cloud-based technologies to build and manage their RAN. This reduces vendor lock-in and enables faster innovation and adaptation to change requirements.

AWS provides various services and solutions, such as Amazon EC2, AWS Outposts, and AWS Wavelength, that can support the deployment and management of O-RAN and vRAN components, making it easier for telecoms to explore and adopt this open interface-based technology.

Implementation steps

- Review the Open Radio Access Network (O-RAN) architecture and understand its key components, such as the O-RAN Radio Unit (O-RU), O-RAN Distributed Unit (O-DU), and O-RAN Central Unit (O-CU).
- Assess the feasibility of adopting an O-RAN and vRAN architecture for your telco network, considering factors like performance, scalability, and cost-effectiveness.
- Evaluate AWS services and solutions that support the deployment and management of O-RAN/vRAN components, such as Amazon EC2, AWS Outposts, and AWS Wavelength.
- Pilot an O-RAN and vRAN implementation in your network, focusing on a specific use case or geographic region to validate the approach.
- Analyze the cost savings and other benefits achieved using open interface-based RAN technologies.
- Based on the pilot results, develop a roadmap for the wider adoption of O-RAN and vRAN across your telco network infrastructure.

Resources

Key AWS services:

- Amazon EC2
- AWS Outposts
- AWS Wavelength

Manage demand and supply resources

TELCOCOST06: How do you prioritize traffic across your hybrid networking connections?

This question focuses on how telcos can effectively prioritize traffic across their hybrid networking connections to optimize resource utilization and costs. In a hybrid networking environment, where traditional on-premises networking elements are combined with cloud-based or virtualized networking components, managing and prioritizing network traffic becomes increasingly complex.

Best practices

- TELCOCOST06-BP01 Implement load-balancing techniques to achieve better utilization of hybrid network resources

TELCOCOST06-BP01 Implement load-balancing techniques to achieve better utilization of hybrid network resources

Establish a comprehensive traffic classification system that identifies different types of network traffic (for example, real-time communications, bulk data transfers, best-effort traffic). Monitor the utilization of your hybrid network connections in real-time. Utilize load-balancing mechanisms to distribute network traffic across multiple hybrid network connections, both on-premises and in the cloud.

Desired outcome:

- Optimize the utilization of networking resources across your hybrid infrastructure.
- Verify critical network traffic is prioritized and routed efficiently.

- Improve overall network performance and cost-effectiveness.

Common anti-patterns:

- Static routing configurations that do not adapt to changing network conditions.
- Lack of visibility into network traffic patterns and resource utilization.
- Inability to dynamically adjust traffic flows based on real-time performance metrics.

Benefits of establishing this best practice:

- Improved networking resource utilization and cost-effectiveness.
- Enhanced performance and reliability for latency-sensitive network traffic.
- Increased agility in managing network capacity and traffic flows.
- Better alignment of network capabilities with evolving business requirements.

Level of risk exposed if this best practice is not established: High

Implementation guidance

Implementing effective load-balancing techniques across your hybrid networking infrastructure is crucial for optimizing resource utilization and cost-effectiveness. By leveraging advanced load-balancing capabilities, you can intelligently route network traffic based on real-time conditions, prioritize critical applications and services, and verify efficient use of your networking resources.

AWS provides various load-balancing services, such as Network Load Balancer and Application Load Balancer, that can assist you to distribute traffic across your hybrid network connections, both on-premises and in the cloud. These load balancers can leverage dynamic routing algorithms to route traffic based on performance metrics like latency, jitter, and packet loss, verifying that traffic is directed to the most optimal path.

Additionally, services like AWS Global Accelerator can be used to optimize the routing of user's voice traffic to the closest available network endpoint, further reducing data transfer costs and improving the user experience.

Implementation steps

- Establish a comprehensive traffic classification system that identifies different types of network traffic (for example, real-time communications, bulk data transfers, best-effort traffic).

- Deploy Network Load Balancer or Application Load Balancer to distribute voice traffic across your hybrid network connections, both on-premises and in the cloud.
- Configure the load balancers to use dynamic routing based on network metrics, such as latency, jitter, and packet loss, to intelligently route traffic.
- Leverage AWS Global Accelerator to optimize the routing of user traffic to the closest available network endpoint, reducing data transfer costs.
- Monitor the utilization of your hybrid network connections in real-time using Amazon CloudWatch and set up alarms to trigger load balancing adjustments.
- Continuously review and refine your load balancing policies to verify optimal utilization of your hybrid network resources.

Resources

Key AWS services:

- AWS Network Load Balancer
- AWS Application Load Balancer
- AWS Global Accelerator
- Amazon CloudWatch

Sustainability

The sustainability pillar focuses on environmental impacts, especially energy consumption and efficiency since they are important levers for architects to inform direct action to reduce resource usage. Environmental sustainability is a shared responsibility between customers and AWS. AWS is responsible for sustainability of the cloud, delivering efficient, shared infrastructure, water stewardship, and sourcing renewable power. Customers are responsible for sustainability in the cloud, optimizing workloads and resource utilization.

Focus areas

- Design principles
- Energy optimization
- Data processing and storage
- Investment protection
- Climate change risk
- Carbon tracking

Design principles

The Well-Architected Framework identifies a set of general design principles to facilitate superior design in the cloud. In addition, the following design principles should also be considered for designing and operating telco workloads:

- **Organization-wide goals:** Organizations should have long-term sustainability goals established based on each workload; types of resources being used for infrastructure and planning for growth. Designing workloads keeping in mind these goals and what future growth gives individual application owners benchmarks to work back from. Telecommunications organizations should keep the following in mind when establishing sustainability goals:
- **Center of sustainability:** Creating a Center of Sustainability within the organization will create a team and a one-threaded leader that can work towards establishing these goals and accelerating achievement of those. They will be responsible for aligning workloads to these goals, reducing current carbon footprint by exploring innovative technologies and optimizing upon existing infrastructure.
- **Network energy efficiency:** Telecommunications Organizations should have guardrails in place to optimize energy consumption of the RAN, transport, core network, and IT infrastructure.

Organizations should also consider developing depreciation schedules for legacy networks that are no longer being utilized and that are not energy efficient.

- **Use O-RAN:** With Open RAN, it is possible for telecom operators to leverage multiple RAN technologies. This gives telecom operators the opportunity to leverage more efficient RAN capabilities from different vendors and suppliers. It allows us to deploy new innovations, like miniaturization of radios, at a much more rapid scale.
- **Monitoring and observability:** Progressing towards sustainability goals is difficult without having visibility across the entire network and IT infrastructure's carbon emissions. Having a robust monitoring system, which provides a single pane of glass on each component's carbon footprint would provide actionable insights that drive sustainability plans. Utilize AWS's carbon footprint measurement tools which provides customers with tools to measure and report on GHG emissions from AWS usage, which complement sustainability frameworks.
- **Renewable energy and green network:** Utilizing renewable energy to power telecom networks can go a long way in reducing carbon emissions. For example, instead of using diesel generators, operators might consider solar power, wind power, lithium-ion batteries, and renewable electricity from a renewable source.
- **Use managed and serverless services:** Enables organizations to automatically scale resources based on actual demand while reducing idle compute time and maximizing resource utilization. This architectural approach shifts compute resources to centralized data centers and leverages cloud efficiencies to significantly reduce power consumption and operating costs while maintaining high availability for telecom operations. The event-driven nature of serverless computing maintains optimal resource allocation by blocking unnecessary resource consumption during low traffic periods and automatic scaling during peak demands.
- **Maximize utilization:** Right-sizing compute, storage, network infrastructure to support an end-to-end telecom network is key to optimizing its footprint. Energy efficient designs to verify high utilization and maximize the energy efficiency of the underlying hardware. Two hosts running at 30% utilization are less efficient than one host running at 60% due to baseline power consumption per host. At the same time, minimize idle resources, processing, and storage to reduce the total energy required to power your workload.
- **Anticipate and adopt new, more efficient hardware and software offerings:** Support the upstream improvements your partners and suppliers make to assist you to reduce the impact of your cloud workloads. Continually monitor and evaluate new, more efficient hardware and software offerings. Design for flexibility to allow for the rapid adoption of new efficient technologies**.**

- **Retire legacy networks (2G/3G) and migrate customers to more energy efficient technologies such as 5G:** Dated hardware and legacy networks lead to more energy consumption and higher carbon footprint. Moving away from them and towards newer technologies can assist CSPs to be more energy efficient.

Energy optimization

TELCOSUS01: How do you optimize energy consumption for the telco Networks?

Optimizing energy consumption in telecommunications networks is critical for reducing operational costs, meeting sustainability targets, and maintaining competitive advantage. This involves implementing energy-efficient infrastructure, intelligent resource management, and continuous optimization of network operations.

Best practices

- TELCOSUS01-BP01 Implement energy-efficient infrastructure for telco networks

TELCOSUS01-BP01 Implement energy-efficient infrastructure for telco networks

When designing telco network infrastructure for sustainability, focus on optimizing compute resources, implementing intelligent scaling, and leveraging edge computing to reduce energy consumption. Start by analyzing your current network functions deployment patterns - including virtual network functions (VNFs), cloud-based network functions (CNFs), and support systems like business support systems (BSS) and operations support systems (OSS).

Desired outcome: Achieve reduction in energy consumption across telco network operations through optimized infrastructure deployment, intelligent auto scaling, and efficient resource utilization while maintaining service availability and meeting service requirements.

Benefits of establishing this practice:

- Reduced energy costs through efficient resource utilization.
- Progress toward net-zero emissions commitments and Science Based Targets.

- Meeting government mandates for carbon reduction in telecommunications.
- Improved performance through right-sized infrastructure.
- Automated scaling reduces manual intervention and human error.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

For compute optimization, use AWS Compute Optimizer to analyze your EC2 instances running telco workloads and receive recommendations for rightsizing. Deploy your network functions on AWS Graviton-based instances which provide up to 60% better energy efficiency compared to comparable x86-based instances. For containerized network functions, use Amazon ECS with AWS Fargate Spot or Amazon EKS with Karpenter to automatically optimize container placement and reduce idle capacity.

Implement edge computing strategies using AWS Outposts, AWS Local Zones, or AWS Wavelength to process data closer to Radio Access Network (RAN) sites and end users. This reduces backhaul traffic and core network energy consumption. For 5G deployments, use AWS Wavelength to embed compute and storage at the network edge within telecommunications providers' datacenters.

Configure AWS Auto Scaling with predictive scaling policies based on historical traffic patterns typical in telco networks (peak hours, special events, seasonal variations). Use Amazon CloudWatch to monitor metrics like CPU utilization, network throughput, and custom metrics from your network functions to trigger scaling actions.

Implementation steps

- Deploy AWS Compute Optimizer and analyze recommendations for EC2 instances running telco workloads. Generate reports to identify over-provisioned instances and potential savings.
- Implement tagging strategy for telco resources using tags like Environment, NetworkFunction, TrafficPattern to enable granular monitoring and optimization.
- Configure AWS Systems Manager to automatically stop non-production telco workloads during off-hours and weekends using Maintenance Windows and Automation documents.
- Deploy Amazon CloudWatch Container Insights for ECS/EKS clusters running containerized network functions to identify idle containers and optimization opportunities.
- Set up AWS Auto Scaling with target tracking policies for core network functions, using metrics like requests per second or concurrent sessions typical in telco workloads.

- Implement AWS Lambda for event-driven telco functions like billing notifications, network alerts, and configuration updates to reduce idle compute.
- Use AWS Cost and Usage Reports with hourly granularity to correlate resource usage with network traffic patterns and identify optimization windows.

Resources

Key AWS services:

- EC2
- Amazon CloudWatch
- EKS
- ECS
- AWS Lambda

Data processing and storage

TELCOSUS02: How do you optimize data processing and storage for sustainable telco operations?

Managing the massive volumes of data generated by telecommunications networks requires efficient lifecycle management strategies to minimize storage costs and energy consumption while maintaining adherence with regulatory requirements and operational needs.

Best practices

- TELCOSUS02-BP01 Implement efficient data lifecycle management for telco networks

TELCOSUS02-BP01 Implement efficient data lifecycle management for telco networks

telco networks generate diverse data types with varying retention requirements. Call detail records (CDRs) requiring long-term retention for regulatory adherence, network performance metrics needed for real-time analysis, and IoT sensor data with high velocity but short relevance periods.

Design your data architecture to automatically optimize storage based on access patterns while maintaining adherence with telecommunications regulations.

Desired outcome: Reduce storage costs and energy consumption through intelligent data tiering, compression, and lifecycle management while maintaining adherence with telecommunications regulatory requirements for data retention and accessibility.

Benefits of establishing this practice:

- Achieve reduction in storage costs through intelligent tiering and compression.
- Automated retention policies meeting telecommunications data regulations.
- Faster data processing through optimized storage patterns.
- Lower energy consumption from reduced storage infrastructure.
- Automated data management reducing manual intervention.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

Start by categorizing your telco data based on regulatory requirements, access frequency, and business value. For CDRs and lawful intercept data, implement immutable storage with defined retention periods. For network telemetry and performance data, use time-series optimized storage with aggressive compression. For customer analytics and billing data, implement tiered storage based on access patterns.

Use Amazon S3 Intelligent-Tiering to automatically move data between frequent, infrequent, and archive access tiers without performance impact or operational overhead. Configure lifecycle policies specific to telco data types (for example, moving CDRs to Glacier Deep Archive after 90 days (about three months) while keeping metadata in Amazon S3 Standard for quick retrieval).

For real-time data processing, implement stream processing architectures using Amazon Kinesis Data Streams for ingesting network telemetry, with Managed Service for Apache Flink for real-time anomaly detection and Firehose for delivering processed data to Amazon S3 in compressed Parquet format.

Implementation steps

1. Create S3 buckets with specific naming conventions for different telco data types, for example: `cdr-data`, `network-telemetry`, `customer-analytics`, or `iot-sensors`.

2. Enable S3 Intelligent-Tiering on S3 buckets and configure archive policies: CDRs to Deep Archive after 90 days (about three months) and network logs to Glacier Instant after 30 days (about a month).
3. Implement AWS Glue ETL jobs to compress and convert raw telco data to columnar formats (like Parquet or ORC).
4. Deploy Amazon Kinesis Data Streams with shard auto scaling for ingesting real-time network data, sized based on peak traffic.
5. Configure Managed Service for Apache Flink applications with SQL queries to detect network anomalies and aggregate metrics in real-time, reducing downstream processing needs.
6. Set up AWS Lambda functions triggered by S3 events to validate data adherence, apply compression, and update metadata in Amazon DynamoDB.
7. Implement Amazon Athena with partition projection for querying historical telco data directly from S3 without maintaining separate data warehouses.
8. Create CloudWatch dashboards to monitor storage metrics, data transfer costs, and lifecycle transition effectiveness.

Resources

Key AWS services:

- Amazon S3
- AWS Glue
- Amazon Kinesis Data Streams/Analytics
- AWS Lambda
- Amazon DynamoDB
- Amazon CloudWatch

Investment protection

TELCOSUS03: How are you protecting investments in the telco network assets?

Maximizing the lifecycle value of telecommunications infrastructure requires implementing circular economy principles, predictive maintenance, and asset optimization strategies to reduce waste and extend equipment lifespan.

Best practices

- TELCOSUS03-BP01 Adopt circular economy principles for telco network assets

TELCOSUS03-BP01 Adopt circular economy principles for telco network assets

Implementing circular economy principles in telco networks requires comprehensive visibility into physical and virtual assets throughout their lifecycle. This includes tracking network equipment from procurement through decommissioning, optimizing hardware utilization to extend lifespan, and establishing processes for equipment refurbishment and responsible recycling.

Desired outcome: Extend network equipment lifespan through predictive maintenance, optimize asset utilization, and establish sustainable end-of-life processes for hardware recycling and repurposing, reducing electronic waste and capital expenditure.

Benefits of establishing this practice:

- Achieve reduction in equipment replacement costs through lifecycle extension.
- Minimized electronic waste through reuse and responsible recycling.
- Improved utilization rates through better visibility.
- Meeting circular economy regulations and corporate ESG commitments.
- Data-driven decisions on asset replacement and maintenance.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

Design an asset tracking system that monitors both physical infrastructure (base stations, routers, switches) and virtual resources (VNF licenses, cloud resources). Use IoT sensors to track equipment health metrics like temperature, power consumption, and performance degradation to predict optimal replacement timing and identify reuse opportunities.

Implement predictive maintenance using machine learning to extend equipment lifespan. By analyzing historical failure patterns and real-time telemetry, you can perform maintenance before failures occur, reducing premature equipment replacement and minimizing electronic waste.

Implementation steps

1. Deploy AWS IoT Core to connect and manage telco network equipment sensors, creating digital twins for major assets like base stations and core network equipment.
2. Configure AWS IoT SiteWise to model your telco asset hierarchy (regions, sites, equipment types) and collect metrics on utilization, energy consumption, and health status.
3. Implement Amazon Timestream to store time-series data from network equipment, enabling long-term trend analysis for predictive maintenance.
4. Create AWS IoT Events detectors to identify equipment approaching end-of-life based on performance degradation patterns specific to telco equipment.
5. Deploy Amazon SageMaker AI to build predictive maintenance models using historical failure data.
6. Set up Quick dashboards displaying asset utilization rates, predicted replacement schedules, and opportunities for equipment redeployment.
7. Implement AWS Systems Manager Inventory to track software licenses and virtual resource allocation, identifying underutilized assets for reallocation.
8. Configure Amazon SNS notifications for equipment lifecycle events (like warranty expiration, maintenance due, and end-of-life) to enable proactive asset management.

Resources

Key AWS services:

- AWS IoT Core
- Quick
- Amazon SNS

Climate change risk

TELCOSUS04: Is your telco network deployment resilient to climate change risks?

Building climate resilience into telecommunications infrastructure is essential for maintaining service continuity as extreme weather events become more frequent and severe due to climate change.

Best practices

- TELCOSUS04-BP01 Enhance telco network resilience to climate change risks

TELCOSUS04-BP01 Enhance telco network resilience to climate change risks

Climate resilience for telco networks requires architecting for extreme weather events, implementing geographic redundancy, and establishing automated failover mechanisms. Design your infrastructure to withstand increasing frequency of floods, heatwaves, storms, and other climate-related disruptions while maintaining service continuity for critical communications.

Desired outcome: Achieve network availability despite climate-related disruptions through geographic redundancy, automated failover mechanisms, and predictive risk management, maintaining continuous service for critical communications during extreme weather events.

Benefits of establishing this practice:

- Maintained network availability during extreme weather events.
- Achieve reduction in climate-related service disruptions.
- Enhanced capacity to support disaster response communications.
- Meeting government requirements for critical infrastructure resilience.
- Improved reputation as reliable service provider during emergencies.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

Start by mapping your network infrastructure against climate risk data to identify vulnerable locations. For high-risk areas, implement additional redundancy and strengthened physical protection. Deploy critical network functions across multiple AWS Regions and Availability Zones to verify geographic diversity. For edge locations, use AWS Outposts with ruggedized options for harsh environments.

Implement real-time environmental monitoring at cell sites and data centers to detect climate-related threats early. Use predictive analytics to anticipate weather-related traffic surges (emergency calls during disasters) and pre-scale resources accordingly. Design automated response systems that can reroute traffic, activate backup sites, and notify operations teams without manual intervention.

Implementation steps

Deploy AWS Outposts or Local Zones in geographically diverse locations, maintaining critical network functions are distant apart to avoid single weather event impact.

1. Implement AWS Backup with cross-Region replication for critical telco workloads, with the appropriate Recovery Time Objectives (RTO) for essential services.
2. Configure Amazon Route 53 health checks with automatic DNS failover for critical endpoints, verifying traffic reroutes upon failure detection.
3. Set up AWS IoT Core to collect environmental data (like temperature, humidity, and water levels) from cell sites, with IoT Analytics to predict climate-related risks.
4. Create AWS Lambda functions triggered by weather API data to automatically scale resources in anticipation of emergency traffic surges.
5. Implement AWS Step Functions to orchestrate complex disaster recovery workflows, including service failover, data synchronization, and stakeholder notifications.
6. Deploy Amazon CloudWatch Synthetics to continuously monitor service availability from multiple geographic locations, simulating user traffic patterns.
7. Configure AWS Systems Manager Automation documents for common climate-related scenarios (like power outage response, flooding protocols, and heatwave procedures).

Resources

Key AWS services:

- AWS Outposts
- Amazon Route 53
- AWSIoT Core
- AWS Lambda
- Amazon CloudWatch

Carbon tracking

TELCOSUS05: How do you measure and track the carbon footprint of your telco network operations?

Establishing comprehensive carbon footprint monitoring is essential for telecommunications organizations to meet sustainability goals, comply with regulations, and demonstrate environmental responsibility to stakeholders.

Best practices

- TELCOSUS05-BP01 Establish comprehensive carbon footprint monitoring for telco networks

TELCOSUS05-BP01 Establish comprehensive carbon footprint monitoring for telco networks

Measuring carbon footprint across telco operations requires tracking emissions from multiple sources including data centers, network equipment, edge sites, and even supply chain activities.

Desired outcome: Achieve complete visibility of carbon emissions across network operations with automated tracking, reporting accuracy, and actionable insights that enable emission reductions aligned with Science Based Targets initiative (SBTi) commitments.

Benefits of establishing this practice:

- Automated reporting for telecommunications environmental regulations.
- Identify and prioritize highest-impact reduction opportunities.
- Accurate sustainability reporting for investors and customers.
- Demonstrated environmental leadership in telecommunications sector.
- Link carbon reduction initiatives to operational cost savings.

Level of risk exposed if this best practice is not established: Low

Implementation guidance

Implement a comprehensive monitoring system that captures the following while providing actionable insights for reduction strategies:

- Scope 1: direct emissions
- Scope 2: purchased electricity
- Scope 3: value chain emissions

Design your monitoring architecture to automatically collect energy consumption data from infrastructure components. Tag AWS resources with sustainability metadata (equipment type, location, and business function) to enable granular carbon accounting. Integrate with AWS Customer Carbon Footprint Tool for baseline cloud emissions data, then augment with custom metrics for on-premises equipment and network operations.

Establish automated reporting pipelines that aggregate carbon data, calculate emissions using regional grid factors, and generate reports for regulatory requirements and sustainability frameworks like SBTi or Global System for Mobile Communications Association (GSMA) climate targets.

Implementation steps

1. Enable AWS Customer Carbon Footprint Tool and configure monthly exports to S3 for historical analysis and trend tracking.
2. Implement comprehensive tagging strategy with tags: `CarbonCategory`, `LocationGrid`, `NetworkFunction`, and `EquipmentType` for AWS resources.
3. Deploy Amazon CloudWatch custom metrics to track energy consumption from on-premises network equipment using AWS IoT Greengrass at edge sites.
4. Create AWS Glue ETL pipelines to combine AWS carbon data with on-premises metrics, calculating total emissions using Regional emission factors.
5. Set up Amazon Managed Grafana dashboards displaying real-time carbon intensity metrics, with drill-down capabilities by service, region, and network function.
6. Configure Quick to generate automated monthly sustainability reports aligned with GSMA or SBTi reporting requirements.
7. Implement Amazon EventBridge rules to trigger alerts when carbon emissions exceed defined thresholds or deviate from reduction targets.

8. Deploy AWS Lambda functions to calculate power usage effectiveness (PUE) for data centers and network efficiency metrics for RAN sites.
9. Use Amazon Forecast to predict future carbon emissions based on network growth projections and planned efficiency improvements.

Resources

Key AWS services:

- Amazon S3
- AWS Glue
- Quick
- Amazon EventBridge
- AWS Lambda

Conclusion

The goal of the Telco Lens for the Well-Architected Framework is to provide architectural best practices for designing and operating reliable, secure, efficient, and cost-effective telecommunications workloads on AWS. In operational excellence, we outline best practices around how people, process and operating models need to be aligned so that workloads running on AWS can support mission critical telecommunication services. Architectures for telecommunications workloads need to incorporate security, and customers need to continuously monitor, measure, test failure, and recovery in the cloud to achieve their desired resiliency and performance objectives. These objectives can be met with significant cost savings by right sizing and establishing governance models around consumption and monitoring of AWS resources

This Framework can improve security, resiliency, and operational efficiency for CSP customers migrating and building applications on AWS and can also assist in meeting regulatory obligations.

This lens provided architectural guidance for designing and building efficient, cost-effective, and scalable telco workloads in the cloud. We captured common architectures and overarching design tenets for key telco workload needs, such as 5G network management, intelligent routing, real-time analytics, and customer experience applications.

The whitepaper discussed the AWS Well-Architected Framework pillars through a telco industry lens, providing a set of questions for you to consider when architecting new or evaluating existing telco workloads. Applying the Well-Architected Framework to your telco architectures assists to build robust, resilient, and scalable cloud-based systems that meet the demanding performance, security, and cost requirements of the telecommunications industry.

Some of the key tenets covered in this lens include:

- Leveraging managed services and serverless architectures to increase operational efficiency and reduce maintenance overhead
- Implementing distributed, fault-tolerant designs to verify high availability and resiliency of mission-critical telco applications
- Automating deployment and operations processes to improve agility and reduce human error
- Optimizing costs through rightsizing, reserved capacity, and intelligent traffic routing
- Embedding security controls from the ground up to protect sensitive customer and network data

By following the guidance in this lens, you can architect telco workloads that are reliable, performant, secure, and cost-effective - assisting your organization to deliver exceptional customer experiences and stay competitive in the fast-paced world of telecommunications.

Contributors

The following individuals and organizations contributed to this document:

- Ammar Latif - Principal Solutions Architect, Amazon Web Services
- Ashish Shah - Senior Solutions Architect, Amazon Web Services
- Aymen Saidi - Principal Solutions Architect, Amazon Web Services
- Christopher Adigun - Senior Solutions Architect, Amazon Web Services
- Hamad Ahsan - Senior Technical Account Manager, Amazon Web Services
- Karunakara Gadde – Principal Technical Account Manager, Amazon Web Services
- Mohamed Daoud - Senior Technical Account Manager, Amazon Web Services
- Naveen Pavuluri - Enterprise Support Manager, Amazon Web Services
- Pablo Forero - Solutions Architect, Amazon Web Services
- Puru Arsikere - Enterprise Support Lead, TAM, Amazon Web Services
- Sarah Morrison - Principal Customer Solutions Manager, Amazon Web Services
- Saurabh Yadav - Enterprise Support Lead, TAM, Amazon Web Services
- Scott Stimson - Principal Solutions Architect, Amazon Web Services
- Scott Wainner - Principal Solutions Architect, Amazon Web Services
- Shirin Bhambhani - Senior Solutions Architect, Amazon Web Services
- Srinivas Kurla - Senior Solutions Architect, Amazon Web Services
- Waqar Afzal - Senior Solutions Architect, Amazon Web Services
- Madhuri Srinivasan, Senior Technical Writer, Amazon Web Services
- Stewart Matzek, Senior Technical Writer, Amazon Web Services
- Matthew Wygant, Senior TPM Guidance, Amazon Web Services

Document revisions

The following table describes the documentation releases for the Telco Lens.

Change	Description	Date
Initial release	Initial release of the Telco Lens.	December 30, 2025

AWS Glossary

For the latest AWS terminology, see the AWS glossary in the *AWS Glossary Reference*.