Bringing Cloud Security to the Open RAN

Abstract

Open RAN allows network operators to source RAN equipment components from multiple vendors, creating a more vibrant ecosystem. As network operators adopt Open RAN, they need cloud security solutions to mitigate the potential risk of the threat surface inherent to open and cloud-based architectures. This whitepaper details the cloud security principles that benefit Open RAN and provides a framework for secure management of Open RAN O-Cloud components.

Bringing Cloud Security to the Open RAN

Introduction

Networking and computing are everywhere today. In the past few years, a large portion of the industry workforce has become remote, and this shift has increased the need for connectivity and compute at the edge. At the same time, the global economy continues to create new scenarios that demand low-latency, high-bandwidth, always-on networking and computing. The proliferation of mobile devices, smart cities, software-defined agriculture, and industrial internet of things (IIoT) has created an unprecedented thirst for connectivity and bandwidth. The telecommunications industry must continue to adopt new technologies to support the growing demands of the global economy.

"The telecommunications industry must continue to adopt new technologies to support the growing demands of the global economy."

Advances in the cloud—both in the core and at the edge of the communications network—allow for the connection of billions of devices. The cloud's immense processing power can be distributed where and when it is needed without sacrificing performance or increasing latency. The cloud can store vast volumes of data which can be accessed from anywhere broadband is available. These cloud characteristics enhance the operational efficiencies available to businesses across every industry. They also make the cloud an ideal technology to support the rise of the IIoT, particularly when combined with the modern wireless network, 5G.

5G's service-based architecture already defines how network functions (NF) can be deployed in the cloud.[1] The following section looks at O-RAN architecture.

Open Radio Access Networks

Open Radio Access Networks or ***Open RAN (O-RAN)*** create open interfaces between all software components. Virtualization[2] and cloudification[3] are already progressing in next-generation wireless networks, and the evolution of O-RAN is an important step toward maximizing the benefits of these efforts. For this reason, most Mobile Network Operators (MNOs) have embraced O-RAN. However, some concerns persist around exposure to an increased threat surface in an open network architecture compared to vendor-specific RAN solutions.

This whitepaper details the cloud security principles that benefit Open RAN and provides a framework for secure management of Open RAN O-Cloud components.

"Open RAN will provide easier interoperability, deeper integration, faster time-to-market, and the ability to integrate the power of big data and artificial intelligence (AI) for a more efficient operation of RAN."

Industry and Open RAN

Open RAN has become a critical topic in political and industrial debates in the United States, the United Kingdom, and the European Union. Technological advances in Open RAN and 5G have generated significant interest from users, policymakers, and the industry for state-of-the-art data flows, better connectivity, and new use cases.

What is clear is that Open RAN and 5G network characteristics are designed to help specific architectural layers improve costs and adaptability. AI will enable further improvements in networks and services, and we already see the industry working on this implementation. Ultimately, Open RAN will provide easier interoperability, deeper integration, faster time-to-market, and the ability to integrate the power of big data and artificial intelligence (AI) for a more efficient operation of RAN.

Figure 1 from 5G Magazine shows the state of Open RAN in 2021. At least 27 countries have Open RAN deployments or are in trial, while 31 operators have active deployments or trials in one or more countries.[4]

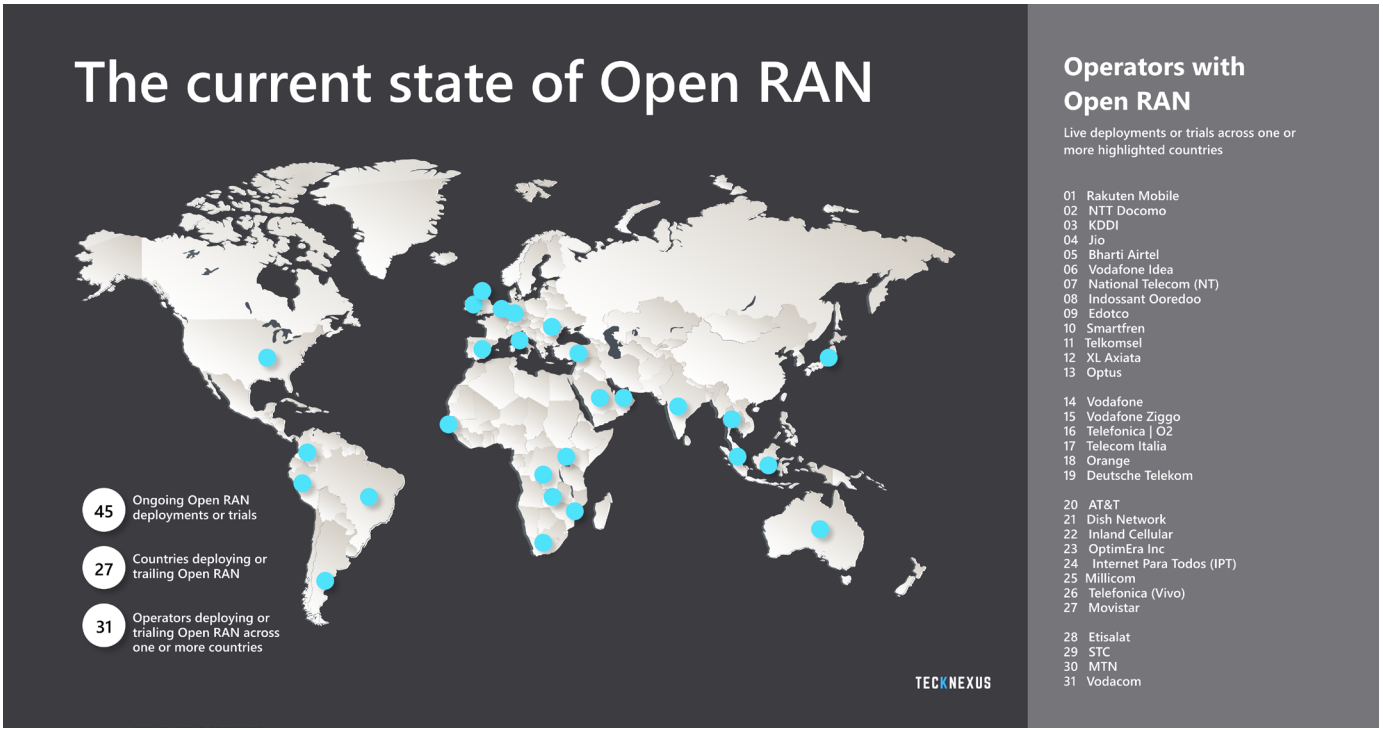


Figure 1: Current state of Open RAN

O-RAN standards and architecture

O-RAN is a new concept that the 3rd Generation Partnership Project (3GPP) has yet to address, causing industry leaders to form the O-RAN Alliance to derive O-RAN technical specifications. While the O-RAN Alliance is not a standards body, the specifications they develop will likely become the guide for standards that will be implemented in the near future.

Figure 2 below shows O-RAN architecture, including the relationship between O-RAN Network Functions, O-RAN service management and orchestrations framework, and the O-Cloud.

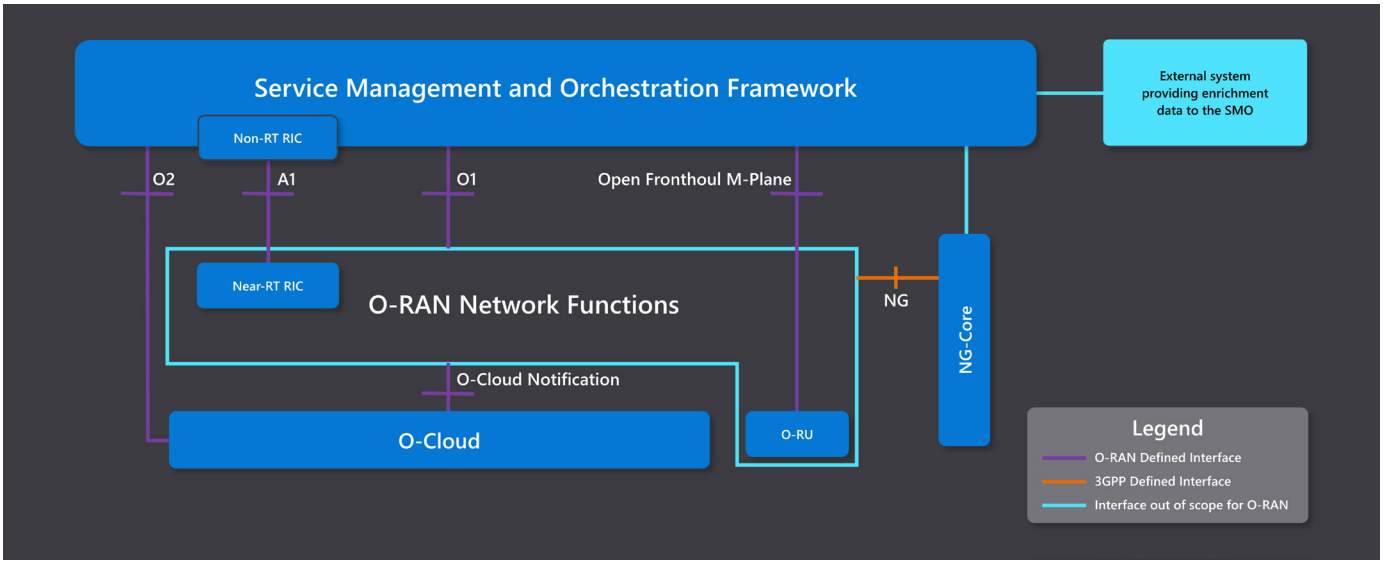


Figure 2: O-RAN high-level architecture[13]

O-RAN network functions deployment

O-RAN network functions are designed and expected to be deployed as depicted in *Figure 3* below. Typical operator deployments include a mix of sites broadly classified as a Radio Unit (RU) and O-RAN Distributed Unit (O-DU)+RU.

Generally speaking, cloud architecture includes edge cloud infrastructure consisting of near-edge and far-edge network functions. Near-edge locations run less-latency-sensitive network functions like the O-CU, Near-RT RIC, etc., and far-edge run latency-sensitive network functions like the O-DU.

The cloud architecture enables a uniform experience regardless of where the workload is deployed, whether far-edge, near-edge, or region. It also extends the cloud services to the edge.

"The cloud architecture enables a uniform experience regardless of where the workload is deployed, whether far-edge, near-edge, or region."

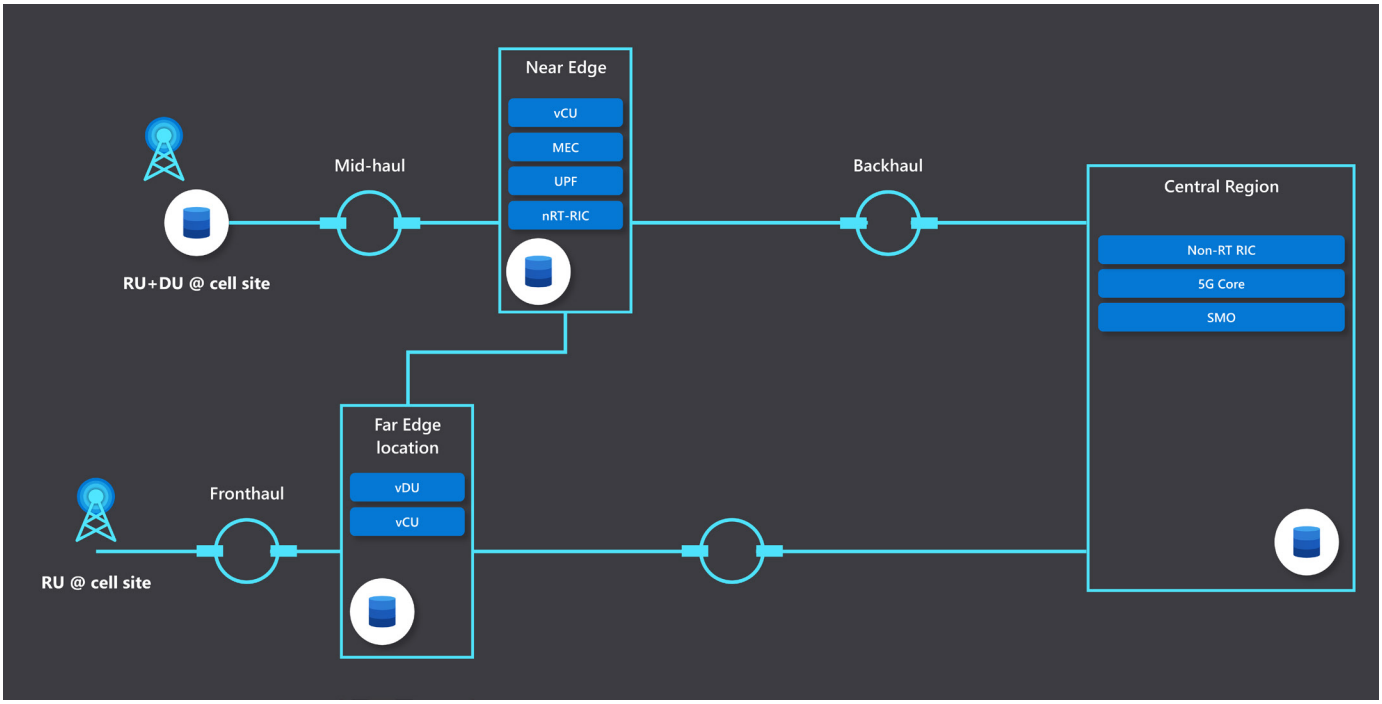


Figure 3: Typical placement of network functions

O-Cloud

According to the O-RAN Alliance, O-Cloud[5] is a cloud-computing platform made up of a collection of physical infrastructure nodes for hosting Open RAN.

This includes:

- Network functions such as Near-RT RIC, O-CU-CP, O-CU-UP, and O-DU
- Software components such as the Operating System, Virtual Machine Monitor, and Container Runtime
- Service management and orchestration functions

Figure 4 below shows the high-level architectural components of the O-Cloud. Some variations may exist between near-edge and far-edge deployments. For example, deployments may utilize accelerators for the far-edge where O-DU workloads require bare-metal performance and for the near-edge where NRT-RIC and O-CU require multi-tenancy.

"O-Cloud is a cloud-computing platform made up of a collection of physical infrastructure nodes for hosting Open RAN."

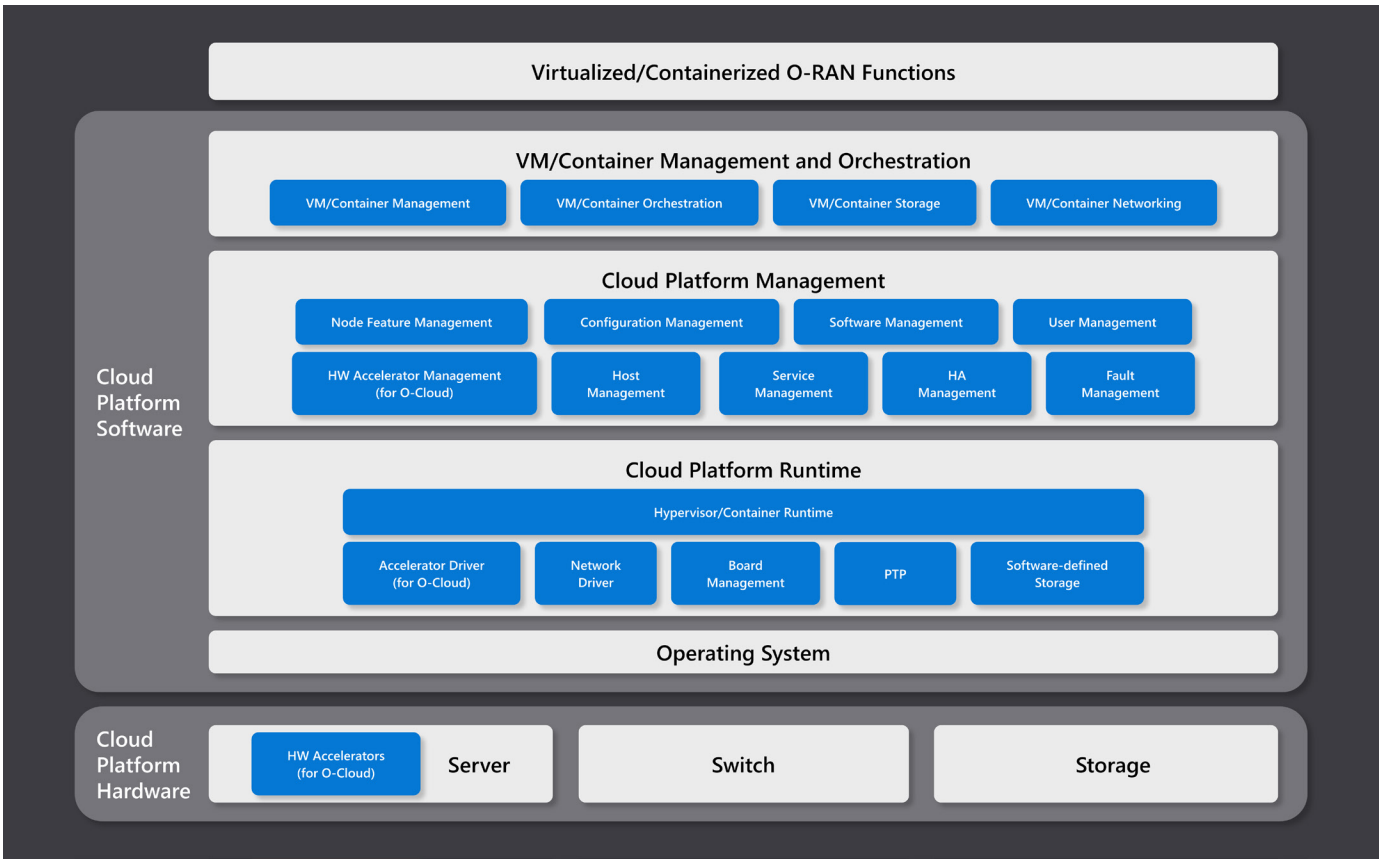


Figure 4: O-Cloud architecture components

Cloud security principles benefit Open RAN

The cloud provides a secure platform for operating the critical infrastructure needs of many sectors across the global economy, including banking, energy, food and agriculture, healthcare, manufacturing, and government services. As such, we see that trust in cloud security is far greater than any other form of computing infrastructure available today. And for a good reason: the cloud remains secure in the face of many vulnerabilities, exploits, and dangers, no matter how big or small. This is exactly why many of the world's largest companies with vital and complex security needs have offloaded much of their network and compute workloads to the cloud.

The cloud's ***massive*** and ***unprecedented*** scale is unique among other computing infrastructure and is precisely what makes the large investments in sophisticated defense and security economically possible. Due to its scale, the cloud's ground-up design includes strict security measures to withstand any type of attack imaginable. The scale of the investment required to defend against sophisticated threats is not logical or economically feasible for small-scale, on-premises systems. This is one of the primary reasons for the supremacy of cloud security over any other computing infrastructure.

O-RAN will benefit from and inherit the cloud's robust security principles when run on the cloud. The inherently modular nature of Open RAN, alongside recent advancements in SDN and network functions virtualization (NFV), unlocks the ability for Microsoft to deploy security capabilities and features at scale across the O-RAN ecosystem. This is accomplished via micro-segmentation and the use of secure containers simultaneously. These characteristics, enhanced by the maturity of Microsoft's industry-leading security, enable more granular control of sensitive data and workloads than prior generations of networking technology. They also allow control in a scalable way through automation and artificial intelligence/ machine learning (AI/ML).

SDN and NFV represent a shift from legacy information and communications technology (ICT) based on hardware to software-based networks capable of leveraging commercial off-the-shelf or commodity-based hardware. This shift rapidly transforms the ICT ecosystem by allowing networks to become more flexible and adaptive while also infusing them with cloud security. Moreover, SDN can facilitate the incorporation or addition of sophisticated security features in real-time by using AI/ML and advanced cloud security capabilities to promptly detect and actively mitigate malicious activities.

The cloud provides significant security benefits when applied to Open RAN. These are organized and discussed in the following sections:

- Secure software development
- Advanced security expertise
- Threat detection and mitigation
- Firmware security
- Supply chain security

Secure software development

Microsoft integrates leading security engineering practices, such as Secure Software Development Standards (SSDS), and operational security capabilities across our suite of infrastructure and services. This includes leveraging the Security Development Lifecycle (SDL) and other software assurance practices in alignment with the National Institute of Standards and Technology's (NIST) Secure Software Development Framework (SSDF)[6] and SAFECode's Fundamental Practices for Secure Software Development to further enhance software security with industry-leading techniques.[7]

Advanced security expertise

Microsoft leverages deep security expertise to monitor and evaluate the security capabilities of our fleet in the face of state-of-the-art security exploits and advanced persistent threats. For example, researchers working in Microsoft's Memory Security Lab continually evaluate the security of Microsoft's cloud in the face of sophisticated threats such as Rowhammer, a security weakness affecting DRAM memory. Similarly, Microsoft's Security Response Center is constantly tracking high-priority, state-sponsored actors to dismantle and preempt their attempts. Such advanced forms of security investments are unheard of in current cellular networks.

Threat detection and mitigation

Each day, Microsoft processes over 8 trillion security signals to assess and protect against threats and attacks that continue to evolve in scale and sophistication. We protect operator networks, services, and infrastructure by using:

- Our massive security signal depth
- A global team of experts
- Our security information and event management (SIEM)
- Our security orchestration, automation, and response (SOAR)
- Detection and response capabilities paired with AI

We integrate this security telemetry with our cloud security capabilities, such as Azure Defender, to block malware and server threats. We also support the broader open-source ecosystem with advanced security through platforms like GitHub and by providing operators access to the Azure Security Center Platform, enabling greater visibility into operator security posture and compliance requirements.

"Each day, Microsoft processes over 8 trillion security signals to assess and protect against threats and attacks that continue to evolve in scale and sophistication."

Firmware security

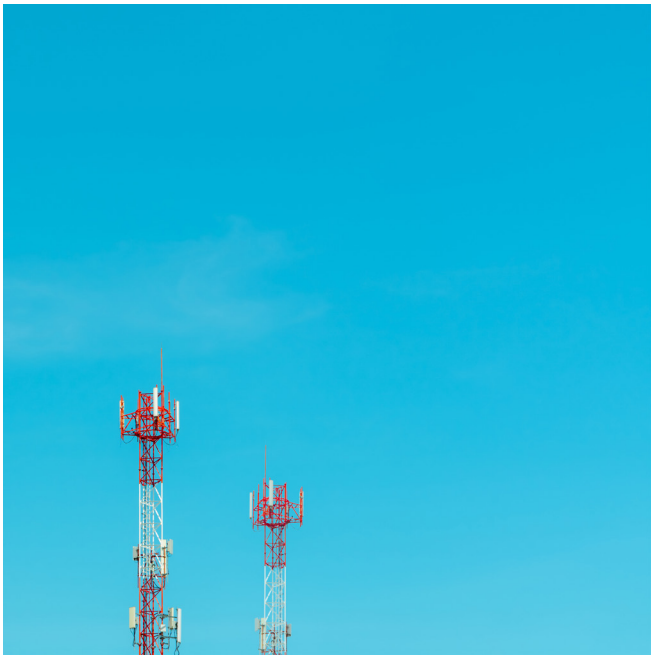
To secure the disaggregated deployments of Open RAN, the Microsoft cloud relies on two roots of trust to verify the integrity of platform firmware: Cerberus and a Trusted Platform Module (TPM).

Cerberus

The first root of trust is Cerberus, a NIST 800-193 compliant hardware root-of-trust with an identity that cannot be cloned.[8] Every Cerberus chip has a unique cryptographic identity established using a signed certificate chain rooted to a Microsoft certificate authority (CA). Cerberus authenticates the integrity of all platform firmware as it is loaded and compares it to the expected value in a platform firmware manifest. If there is a mismatch, the code is not executed, and a remediation process is initiated to restore the server to a trusted state.

Trusted platform module

The second root of trust is the TPM. The TPM is a standardized security component on modern commodity servers with firmware supplied by a trusted third party. The TPM is a secure coprocessor—a dedicated tamper-resistant chip located on the system's motherboard—that provides platform security features in hardware through integrated cryptographic keys. Today, TPMs are used on Linux servers to provide security functions like a *measured boot, remote attestation*, and *disk encryption*.



"Microsoft's two roots of trust deliver a high standard of trust and security to traditional hardware devices—allowing them to be deployed confidently in difficult or insecure physical environments."

Measured boot is an integrity protection mechanism. It ensures that each component in the system boot chain—for example, the Unified Extensible Firmware Interface (UEFI), BIOS, the bootloader, etc.—is measured and that those measurements are securely recorded in the TPM. Measured boot ensures that system configuration measurements match the expected value before allowing the system to complete its boot process.

Remote attestation takes the measurement process one step further—the TPM provides a digitally signed attestation message that contains certain TPM measurements. This attestation can be provided to a remote service, continuously validating the system configuration. The Linux kernel integrity measurement architecture (IMA) subsystem uses the TPM to generate attestation messages for the current kernel measurement list. The remote service can then validate the attestation messages and verify that those measurements came from that specific machine.

During boot, the TPM measures and records firmware and configuration settings for all system components. The boot measurements are cryptographically signed by the TPM and sent to an Azure Host Attestation Service for validation. The measurements must pass validation before the platform is granted permission to join the Azure fleet and host customer workloads. If the system detects a measurement mismatch, the transaction is considered invalid, and the servers are taken offline to be reimaged and returned to a compliant and trusted state.

Microsoft's two roots of trust ensure a robust level of security is integrated into the firmware that sits atop 5G-deployed hardware. Ultimately, this delivers a high standard of trust and security to traditional hardware devices—allowing them to be deployed confidently in difficult or insecure physical environments.

Supply chain security

Securing the global supply chain of information and communications technology (ICT) is more complex than ever. Protection against persistent and increasingly sophisticated threats to both hardware and software is critical to achieving supply chain security. Protecting and strengthening the supply chains for U.S. software and hardware demands a mixture of new and old techniques and methods that enhance the resiliency of information technology (IT) infrastructure.[9]

At Microsoft, we integrate leading supply chain security risk management practices to enhance the security of cloud services deployed in an O-RAN. This is accomplished by leveraging NIST's 800-161 *Supply Chain Risk Management Practices for Federal Information Systems and Organizations*.[10] We also partner with NIST's National Cybersecurity Center of Excellence to enhance the broader software security supply chain through industry-driven proofs-of-concept in secure development automation technologies and practices. Our partnership supports NIST's goal of accelerating and scaling secure software development technologies.[11]

In addition to the cloud-native capabilities that enhance the security of supply chains, we also encourage regulators to find ways to partner across government agencies to incentivize digital solutions in O-RAN to further mitigate supply chain risk. For example, software security technologies designed into software packages or code can address security risks by ensuring trust and preventing software from being exploited by bad actors and malignant uses. These features include:

- The ability to deploy trusted software updates, including the firmware of compromised devices
- Automating security policies to, for example, seek out and prevent user or administrator credentials placement in software code
- In appropriate cases, once in-development standards are finalized, using software bill of materials (SBOMs) to convey evidence that software consumers can trust the environment in which software was built

Similarly, security technologies built into hardware can further protect against supply chain risks. Solutions include hardware roots-of-trust to verify, protect, or restore system, data, or code integrity. They also include secure coprocessors for more robust identity verification and, in appropriate cases, origin and identity attestation for components in a hardware system. For more sensitive applications, data security technologies can protect the exposure of U.S. data through the supply chain. Features include digital rights management, information flow controls, data tagging, and, where appropriate, using secure virtual or data lockbox environments.

Securing Open RAN components

In this section, we explore how existing cloud security practices and tools can be leveraged to secure Open RAN components, including:

- Hardware such as servers and accelerators (as applicable)
- Operating systems
- Cloud runtime as managed Kubernetes and Virtual machine
- Switch fabric

Hardware

There are numerous aspects of the O-Cloud hardware platform that impact security. In this section, we discuss four elements of central importance to O-RAN and O-Cloud security at the hardware platform layer:

- Base platform security
- Isolation of I/O devices
- Key protection
- Confidential computing

Base platform security

Base platform security consists of the hardware, firmware, and software to ensure the integrity of platform components responsible for system operation. O-RAN and O-Cloud benefit from the significant investment in hardware and firmware security technologies that have been developed and are now widely deployed in commodity x86 server platforms.

The key features in today's commodity servers are:

- UEFI secure boot
- A hardware TPM
- A hardware root of trust for firmware updates.
- Low-level system remote access is typically provided by the *Baseboard Management Controller (BMC)*

The **BMC** provides a key component of security for commodity servers. It is typically the lowest-level hardware root of trust in a server platform. The BMC implements the mechanism for local and remote firmware upgrades and performs digital signature validation of all firmware images provided by the system operator to the BMC upgrade mechanism. BMCs also provide the ability for system administrators to remotely access and manage the system console, including power management actions, such as the ability to initiate remote reset and reboot. While BMC implementations across different server original equipment manufacturers (OEM) are not standardized, the APIs for utilizing BMC are standardized with the Redfish API. Examples of the BMC for major OEM server vendors include Dell's iDRAC system and Hewlett-Packard Enterprise's Integrated Lights-Out (iLO) system.

UEFI secure boot detects tampering with bootloaders, key operating system files, and unauthorized firmware

on I/O devices by validating their digital signatures. As part of the boot process, the UEFI firmware checks the cryptographic signature of each piece of the boot software, including UEFI I/O device firmware (also known as option ROMs), EFI applications, and the operating system (including the software bootloader, or GRUB, the operating kernel, and all loadable kernel modules). If all the signatures are valid, the server boots, and the firmware gives control to the operating system. The intent is to prevent malware from modifying the low-level system components listed above.

Isolation of I/O devices

Compared to other virtual network functions (VNFs) and cloud-native network functions (CNFs) used in the telco infrastructure stack, O-RAN network functions are extremely compute-intensive RAN workloads. This is especially important at the lowest layer of the RAN stack, the physical (PHY) layer. As a result, workload-specific hardware accelerators are required for most O-RAN installations.

Although typical O-RAN installations are not *multi-tenant* (the operator is the tenant), the CNFs and VNFs on one operator's infrastructure are often provided by multiple software vendors. From an operator's perspective, the aim is to ensure security and fault isolation among these different network functions—even those functions that access hardware acceleration devices. To provide isolation, O-Cloud relies on Single Root-I/O virtualization (SR-IOV).

SR-IOV allows devices to be shared in a virtualized environment while maintaining protection without taking a significant performance penalty. It also enables a single physical (PCIe) device to appear as multiple separate virtual PCIe devices. Physical devices expose multiple virtual functions (VF) on a single physical server. The hypervisor can assign each VF to a different virtual machine (VM) or to a specific container in a cloud-native environment. VFs ensure the VMs and containers can only access the data and resources on the device that is associated with their own VF.

Key protection

Securing customer keys is a significant challenge, especially in cloud and edge environments. Recent Intel platforms support *Key Locker*, a technology that

allows for data encryption and decryption without direct access to the encryption key. Programs that want to use keys for encryption and decryption operations will instead use *key handles*.

Key handles differ from raw encryption keys in that the keys are no longer directly accessible in memory. Instead, they are bound to a particular system and can also be revoked. After a key handle has been created from an original key, the original key can be erased from system memory without affecting a program's use of the key for encryption and decryption. This ensures that an adversary, even one that has already compromised the system, cannot obtain the original key except during the brief window of time from when the key is created to when it is converted into a handle.

To further address the limited window of opportunity for attackers to obtain the key if they have already compromised the system at the time of key creation, Intel enables Key Locker to be used from within an Intel® Software Guard Extensions (Intel SGX) enclave. Moving key creation into an Intel SGX enclave provides integrity and confidentiality protection. This ensures that an attacker cannot obtain the key during the brief time window between key creation and erasing it after it has been converted into a key handle.

The security benefit of Key Locker is the prevention of an attacker who breaks into a system from being able to use wrapped keys after revocation, on a different system, or in violation of specified key handle usage restrictions (such as using a key handle at ring 3 for a handle that specified only ring 0 usage at handle creation time). Key Locker also offers significant resistance to side-channel attacks.

Key Locker and SGX are only available on recent Intel server platforms and cannot be used on AMD or ARM commodity servers.

Confidential computing

Confidential computing protects data in use by performing computations in a hardware-based Trusted Execution Environment (TEE). TEE environments are secure and isolated to prevent unauthorized access or modification of applications and data while in use. The result is increased security assurances for organizations that manage sensitive and regulated data.

The different commodity server platforms available today offer other methods for confidential computing depending on the CPU vendor and which generation of CPU is being used. Major CPU vendors such as Intel, AMD, and ARM offer different confidential computing technologies, and each includes its security guarantees.

While O-RAN and O-Cloud specifications do not require confidential computing, the pre-existing hardware support for TEEs will make it straightforward for VNFs and CNFs running on O-Cloud platforms to leverage the security benefits of confidential computing. As the technology matures and NF vendors gain experience using confidential computing, future versions of the O-Cloud standards could require it. Below, we summarize the different TEE technologies available today and in the near future, along with a comparison of the differences in their security guarantees and the other limitations that may affect their usage by VNFs and CNFs.

Software

Operating system

CBL-Mariner (CBL) is a Linux distribution for Microsoft's cloud infrastructure and edge products and services. CBL hosts the BareMetal and Container operating system (OS) for the Azure Operator Distributed Services (AODS) platform.

CBL's design comes from the idea that a small, common core set of packages can address the universal needs of cloud and edge services while allowing operators to add packages on top to produce workload images.

CBL's common build system enables:

- Package generation to produce the desired set of RPM packages from SPEC and source files
- Image generation to deliver desired image artifacts from a given set of packages, such as ISOs or VHDs
- End-to-end security and trust based on a cryptographic guarantee of components from source to grave, Federal Information Processing Standards (FIPS) certification, secure enclaves, and access control policies
- Compliance scanning to create a unified pipeline of security and compliance for ingestion, scanning, monitoring, and alerting
- Lifecycle management, including live kernel patching

Whether deployed as a container or a container host, CBLs security benefits include:

- Threat surface reduction
- Supply chain assurances
- Limited consumption of disk and memory resources
- Base image size reduction of eighty percent
- Forty percent reduction in node image upgrade times

The build system's lightweight characteristics and a secure-by-default design made it a logical choice for the AODS platform.

Protecting workloads from internal attacks

Access management for cloud resources is a critical function for any organization using the cloud. Microsoft recommends implementing access management policies in the O-Cloud RAN to protect the platform's workloads from internal attacks, known as cross-container intrusions. Azure role-based access control (Azure RBAC) helps manage who has access to Azure resources, what they can do with those resources, and what areas they can access.

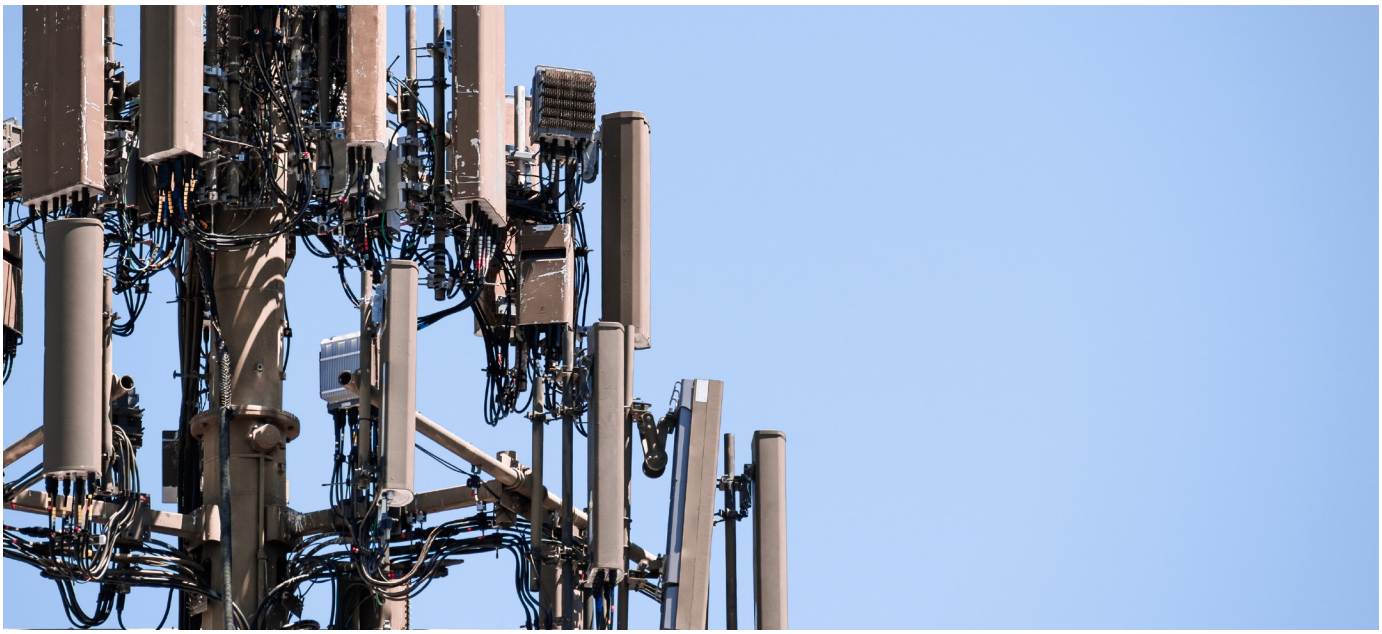
AODS deploys Azure Kubernetes Service (AKS) clusters with Azure Active Directory (Azure AD) integration. Using Azure AD centralizes the identity management component, ensuring that any change in a user account or group status is automatically updated in access to the AKS cluster.

Container and container image security is a significant priority while developing and running applications in AKS. Containers with outdated base images or unpatched application runtimes introduce security risks and possible attack vectors. Container images are scanned for vulnerabilities before any deployment. Only validated images are deployed. Base images and application runtime are updated regularly to redeploy workloads in the AKS cluster.

Switch fabric providing connectivity

The switch fabric is a combination of hardware and software that controls traffic to and from a network node using multiple switches. The topology can be quite complex, and it requires the application of several techniques, including:

- Network isolation
- Virtual networks
- Virtual private networks (VPNs)
- Intrusion detection
- Vulnerability scanning



Network isolation

Network isolation prevents unwanted tenant-to-tenant communications, and access controls block unauthorized users from the network. Virtual machines do not receive inbound traffic from the internet unless customers configure them specifically to do so. For example, the overarching principle within Azure is to allow only connections and communications that are necessary for cloud services to operate, blocking all other ports and connections by default.

- The fabric manages the controller
- The controller provides policy to the switch
- The switch encapsulates the traffic
- Each tunnel has a unique key per VNet
- Policy ensures only good packets get to the proper Azure resource (VM)

Virtual network

A customer can choose to assign multiple deployments within a subscription to a virtual network and allow those deployments to communicate with each other through private IP addresses. Each virtual network is isolated from other virtual networks.

VPN

It connects customers and remote sites to a virtual network using site-to-site and point-to-site VPNs.

Intrusion detection

- Auditing and certification
- Live site penetration testing
- Centralized logging and monitoring
- Fraud and abuse detection

Vulnerability scanning

Multiple layers of automatically updated anti-virus protection are used to protect malicious code from entering the environment. Intrusion detection and prevention systems are in place to detect, alert, and (where applicable) prevent anomalous activities or deviations from a baseline configuration that may indicate a suspected compromise.

Adopt zero-trust security

Microsoft recommends that operators institute a zero-trust (ZT) security model that minimizes breaches and works to prevent damage—even when an employee's identity or an organization's network has been compromised.

ZT is an "assume breach" security posture that treats each step across the network as a unique risk to be evaluated and verified. Instead of assuming everything behind a firewall is safe, the ZT model assumes breach and verifies each request as though it originates from an open network. Regardless of where the request originates or what resource it accesses, ZT requires organizations to "never trust, always verify."

Microsoft enables its users to implement ZT across their hybrid networks by focusing on three core tenants: *Verify explicitly, use least privilege access,* and *adopt an assume-breach mindset*.

Verify explicitly

Verify explicitly policies should always make security decisions using all available data points, including identity, location, device health, resource, data classification, and anomalies. In recent years, *verify explicitly* has expanded to include verifying across your software supply chain. In a ZT model, every access request is fully authenticated (via MFA), authorized (via least privilege access and other frameworks), and encrypted before access is granted. The same security policies are applied regardless of whether the device is government or personally owned and whether it is fully managed by IT or only the apps and data are secured. This applies to all endpoints wherever they are connected. This strategy is compelling because it increases agility, accelerates the speed of detecting threats, and improves an organization's ability to manage the Internet of Things (IoT) and Operational Technology (OT)—while enhancing the end-user experience.

Least privilege access

Use **least privilege access** to limit access with just-in-time and just-enough-access (JIT/JEA) and risk-based adaptive policies. ZT deployments apply least privilege access to infrastructure, ensuring compartmentalized access to systems that can add or modify permissions or policies. This process is no different when applied to hybrid system architecture. Least privilege access reduces risks by minimizing horizontal movement in the event of a breach. ZT leverages least privilege access principles to control verified internal access—a system that restricts access privileges to only those who require it for specific jobs, rather than granting broad access to verified users. The least privilege access principles limit user access with JIT/JEA, risk-based adaptive policies, and data protection to help secure data and prevent attackers from spreading across the network. ZT also uses other variables to determine whether to grant access. Variables used to evaluate access requests include user identity, the request's context, and the access environment's risk.

"Microsoft enables its users to implement ZT across their hybrid networks by focusing on three core tenants: Verify explicitly, use least privilege access, and adopt an assume-breach mindset."

Assume-breach mindset

Finally, **adopt an assume-breach mindset** where the goal is to minimize the blast radius of an incident by leveraging micro-segmentation, end-to-end encryption, continuous monitoring, and automated threat detection and response.

Microsoft recommends operators consider several cyber hygiene practices to improve their security postures towards ZT. A robust ZT approach to security begins with multi-factor authentication (MFA),[12] a process in which users are prompted to provide an additional form of identification during the sign-in process. This additional code could include a code sent to the user's mobile device or a fingerprint scan.

Most cyber-attacks globally are related to passwords. However, governments can protect against 99.9 percent of identity attacks by requiring strong authentication. Governments should always authenticate and authorize users based on all available data points, including user identity, location, device health, service or workload, data classification, and anomalies. Administering MFA is inexpensive and easily incorporated into cloud-based infrastructure. Within cloud services, providers can easily enable MFA to prompt users and groups for additional verification during sign-in.

O-RAN assets identified in the security focus group (SFG) are shown in *Table 1* below.

Thread ID	Risk description	Impact An attacker compromises VNF/CNF images and embedded secrets description	CIA	Severity	Affected assets	Affected components
T-OCloud-01	An attacker compromises VNF/CNF images and embedded secrets	Intellectual property loss and exposure of significant technical details on O-RAN VNF/CNF image. If a registry is comprised it can cascade to downstream VMs, CNs, and hosts.	C, I, A	**HIGH**	D-15, D-16, D-17, D-18, D-20, D-23	O-CLOUD
T-OCloud-02	An attacker exploits weak orchestrator configuration, access control and isolation	Unauthorized hosts join O-Cloud infrastructure and run VMs/CNs; If a single VM/CN host is compromised it can cascade to all VMs/CNs. If authentication key pairs are shared across all VMs/CNs; orchestrator, administrator, and host communications may be authenticated and unencrypted.	C, I, A	**HIGH**	D-12, D-13, D-14, D-15, D-16, D-17, D-18, D-19, D-20, D-24, D-29	ALL
T-OCloud-03	Misuse of a VM/CN to attack other VM/CN, hyper vision/container engine, other hosts (memory, network, storage)	O-RAN system performance degradation, service disruptions, depriving required resources of various O-RAN operating functions.	A	**HIGH**	D-12, D-13, D-14, D-15, D-16, D-17, D-18, D-19, D-20, D-24, D-29	ALL
T-OCloud-04	Spoofing and eavesdropping on network traffic	If VMs/CNs are given direct access to the underlying network stack. It allows other VMs/CNs to intercept and spoof network traffic destined for co-hosted VMs/CNs. Secondly, direct access to the underlying network would allow attackers to gain valuable information on the internal network traffic.	C, I	**HIGH**	D-12, D-13, D-14, D-15, D-16, D-17, D-18, D-19, D-20, D-24, D-29	ALL
T-OCloud-05	An attacker compromises auxiliary/supporting network and security services	Unauthenticated/unauthorized access of the auxiliary/supporting VM/CN leads to compromised O-Cloud, lateral attack towards O-RAN system, component(s) from inside, and potential loss/stolen/tampering of sensitive data.	C, I	**HIGH**	D-12, D-13, D-14, D-15, D-16, D-17, D-18, D-19, D-20, D-24, D-29	ALL

Table 1: O-Cloud threat inventory

Identifying security threats

Leveraging the cloud enables our customers to deploy specific management capabilities, such as Azure Sentinel, for secure capabilities across transport layer security (TLS), role-based access controls (RBAC), authentication policies, and more.

Azure Sentinel is a scalable, cloud-native, security information event management (SIEM) and security orchestration automated response (SOAR) solution. It delivers intelligent security analytics and threat intelligence across an operator's enterprise and provides a single solution for alert detection, threat visibility, proactive hunting, and threat response. In particular, Sentinel's capabilities allow our customers to deploy security automation and orchestration, unlock investigative tools, leverage threat hunting, and collect and analyze data.

Another critical component of our security engineering practices is testing. At Microsoft, we rigorously test our products and services during the development, deployment, operations, and maintenance phases.

The three primary testing practices we integrate are:

1. Static analysis security testing (SAST)
2. Dynamic analysis security testing (DAST)
3. Penetration testing

SAST

SAST provides a highly scalable method of security code review. This is accomplished by analyzing source code before compilation to validate the use of secure coding policies.

DAST

DAST enables the performance of runtime verification of fully compiled or packaged software. It checks the functionality of end-to-end security and empowers the consistency of all integrated components. This is typically achieved using a tool, suite of prebuilt attacks, or tools that precisely monitor application behavior for memory corruption, user privilege issues, and other critical security problems.

Penetration testing

We employ the ZT "Assume Breach" strategy in testing as well. We do this by using highly specialized groups of security experts, known as Red Teams, to perform offensive activities. We use Blue Teams to improve defenses. The teams strengthen enterprise cloud services' threat detection, response, and defense.

While these practices have been in place for years, most customers are unaware of the work behind the scenes to harden the Microsoft enterprise cloud. We run Red Team penetration testing to uncover potential vulnerabilities resulting from coding errors, system configuration faults, or other operational deployment weaknesses. Our Red Teams are made up of skilled security professionals simulating the actions of a hacker. The tests are often performed in conjunction with automated and manual code reviews to provide greater analysis than would ordinarily be possible.

Learn more: Microsoft Enterprise Cloud Red Teaming

Use cases

In this section, we look at two use cases that demonstrate how cloud security solutions protect an operator's Open RAN.

Critical vulnerability patching

Scenario

A proof of concept (POC) for a ubiquitous software package is publicly disclosed, and an exploit code is immediately made available. The exploit code facilitated remote code execution capabilities within the context of the user running an application that leverages the vulnerable software package. The corporate security team receives signaling to immediately begin an assessment of the impact on the network functions virtualization infrastructure (NFVi) that supports sensitive workloads in 100+ production regions. The following day, the vulnerability is rated a 10 out of 10 on the common vulnerability scoring system (CVSS). Global fallout from the exposed vulnerability ensued.

Identification

After early warning signals, the security team immediately started to assess the vulnerability, its potential impact on the environment, and its applicability. This vulnerability presented a dimension of complexity in the software that could exist in multiple forms.

Although the software stack was large, the security team could quickly assess what and how many versions of the platform (NFVi) were in production regions. They did this by leveraging a declarative deployment structure with a mature versioning structure. By knowing the assessment inventory, the team could prioritize software analysis based on the threat surface. Adopting cloud-native principles allowed the organization an SBOM for each deployed version of the platform. This assisted in understanding the components, versions, dependencies, and source code within each release. The ability to instantly execute targeted security analysis uncovered components of interest that required immediate attention.

Response

The team commenced an effort to validate exploitability by using the knowledge of the impacted platforms. Then, they executed as an additional act of due diligence because environment variables for O-Cloud deployments may vary. For this reason, it is critical to understand the differentiation between the existence of a vulnerable piece of software and the ability to exploit a vulnerable piece of software before issuing a mandate for remediation.

The team could not successfully exploit the vulnerable software because of the defense-in-depth posturing of the platform. However, given the high severity and visibility of the vulnerability, the team moved forward with delivering remediation guidance to the appropriate development teams.

Post-mortem

The foundational principles of cloud-native infrastructure saved the security team enormous time and effort by quickly and accurately determining the entire asset inventory using an SBOM. The team was able to focus its energy on validation and remediation coordination. In a span of 48 hours, the security team was able to:

- Identify how many production regions were impacted
- Identify how many individual units of software were impacted
- Actively test against known impacted platform versions
- Prioritize remediation efforts against known exploit paths
- Deliver actionable guidance to the developers
- Provide timely reporting to leadership

The delivery of a remediated platform version for lab testing took less than seven days.

Predictable infrastructure delivery

Challenge

As telco operators look to cloud-native capabilities, they must be able to predictably deliver raw infrastructure while simultaneously having the means to facilitate lifecycle management functions at a reasonable pace without sacrificing availability or integrity. Considering the often-heterogeneous makeup of environments and workloads, the task of finding a single solution to meet all requirements is unlikely.

Solution: Azure Operator Distributed Services

The Azure Operator Distributed Services (AODS) is a carrier-grade hybrid cloud service designed to meet operators' needs by facilitating both the deployment and management of operator-owned, on-premises infrastructure. AODS utilizes object-oriented resource modeling for the provisioning and life-cycling of on-premises infrastructure, including bare-metal, network fabric, and virtual infrastructure.

AODS provides interfaces and the required connective tissue to instantiate on-premises infrastructure in this approach. In this way, AODS supports the runtime and orchestration of network functions such that the operator must only express intent through manifests. Then, AODS consumes and actions against the manifest.

There are two primary objectives to this approach:

- To produce and deliver a rich Azure experience to telco operators
- To integrate with the existing Azure and Arc-connected ecosystem, provide a one-touch cloud deployment experience, free operators to focus on business objectives instead of managing and financing the overhead of human-centric operations

Key design principles

AODS provides the capability for predictable infrastructure delivery by adhering to a set of key design principles which meet the demands of operators' high-performing workloads.

The AODS key design principles are as follows:

- **High availability**: No single point of workload failure.
- **High capacity and scalability**: Uniform data plane and control plane scale. Support the scaling of forwarding tables, the scaling of network segments, Layer 2 segment extension, virtual device mobility, forwarding path optimization, and virtualized networks for multi-tenant support on shared physical infrastructure.
- **Cost and power efficiency: Off-the-shelf hardware with several power profiles** to ensure applicability to all data center power requirements.
- **Simplicity**: Leaf Spine Architecture with EVPN/ VXLAN, predictable traffic flow, high-bandwidth, low-latency, nonblocking server-to-server connectivity.
- **Modularity**: Uses the same hardware and rack design for all BOM options. Integrators may pre-configure a rack before shipment to the operator's location.
- **Flexibility**: Allows operators to choose a variety of instance sizes and with a variety of uplink bandwidth speeds.
- **Non-disruptive expansion**: Allows the introduction of net-new racks without impacting existing racks or re-cabling.
- **Longevity**: Operator investment in on-premises infrastructure needs to support a lifespan of 5 years. It must support performance improvements in the virtualized network functions during this time without demanding new physical deployments.

Conclusions

Microsoft is helping advance carrier-grade edge-cloud solutions that empower operators globally to deploy Open RAN network functions easily and securely. Our tools and services can manage RAN deployments at scale. With Azure AI/ML, core components of our technologies, operators can perform analytics that optimize performance, improve management, and proactively detect and solve problems.

Security principles designed for the cloud and put in practice over decades are being adopted to make the platform resilient, to prevent, detect, and respond to threats in the network and across the firmware and telecommunications supply chains. Edge and network monitoring and programmability via open APIs will enable a new generation of 5G applications while simultaneously improving operational efficiency. Operators can increase revenues and reduce infrastructure costs while building future-proof solutions.

As we build on our promise, we encourage operators and ecosystem partners to contact us to learn more.

“Microsoft is helping advance carrier-grade edge-cloud solutions that empower operators globally to deploy Open RAN network functions easily and securely.”

Legend of abbreviations

3GPP	3rd Generation Partnership Project
ICT	Information and Communications Technology
IIoT	Industrial Internet of Things
5G	Fifth Generation of cellular technology
AI	Artificial Intelligence
AI/ML	Artificial Intelligence/Machine Learning
AKS	Azure Kubernetes Service
AMD	Advanced Micro Devices, Inc.
AODS	Azure Operator Distributed Services
API	Application Programming Interface
ARM	Advanced RISC Machines
Azure AD	Azure Active Directory
Azure RBAC	Azure role-based access control
BIOS	Basic Input/Output System
BMC	Baseboard Management Controller
CA	Certificate Authority
CBL	CBL-Mariner
CIA	Confidentiality, Integrity, and Availability
CN	Cognitive Network
CNF	Cloud-native Network Function
CPU	Central Processing Unit
CVSS	Common Vulnerability Scoring System
DAST	Dynamic Analysis Security Testing
EVPN/VXLAN	Ethernet VPN/Virtual Extensible LAN
FIPS	Federal Information Processing Standards
GRUB	GRand Unified Bootloader
ICT	Information and Communications Technology
iDRAC	Integrated Dell Remote Access Controller
ILO	Integrated Lights-Out
IMA	Integrity Measurement Architecture
Intel SGX	Intel® Software Guard Extensions
I/O	Input/Output
IoT	Internet of Things
IIoT	Industrial Internet of Things
ISO	International Organization for Standardization
IT	Information Technology
JIT/JEA	Just-in-Time and Just-Enough-Access
MFA	Multi-Factor Authentication
ML	Machine Learning
MNO	Mobile Network Operators
NF	Network functions
NFV	Network Functions Virtualization
NFVi	Network Functions Virtualization infrastructure
NIST	National Institute of Standards and Technology
NRT-RIC	Near-Real-Time Remote Intelligent Communications
O-CU	O-RAN central unit
O-CU-CP	O-RAN Central Unit – Control Plane
O-CU-UP	O-RAN Central Unit – User Plane
O-DU	O-RAN Distributed Unit
OEM	Original Equipment Manufacturers
O-RAN	Open Radio Access Networks or Open RAN
OS	Operating System
OT	Operational Technology
PHY	Physical Layer
POC	Proof of Concept
RAN	Radio Access Network
RBAC	Role-Based Access Controls
ROM	Read-Only Memory
RU	Radio Unit
SAST	Static Analysis Security Testing
SBOM	Software Bills of Material
SDL	Security Development Lifecycle
SDN	Software Defined Network
SFG	Security Focus Group
SOAR	Security Orchestration Automated Response
SR-IOV	Single Root-I/O Virtualization
SSDS	Secure Software Development Standards
TEE	Trusted Execution Environment
TLS	Transport Layer Security
TPM	Trusted Platform Module
UEFI	Unified Extensible Firmware Interface
VF	Virtual Functions
VM	Virtual Machine
VNF	Virtual Network Function
VPN	Virtual Private Network
ZT	Zero-trust

Endnotes

1. Microsoft. *What is cloud computing? A beginners guide.* Accessed June 2022.

2. Microsoft. *What is virtualization?* Accessed June 2022.

3. Microsoft. *What is cloud computing? A beginners guide.* Accessed June 2022.

4. TeckNexus, 5G Networks Magazine. *Current State of Open RAN. Where & who is deploying or trialing Open RAN?* June 2021.

5. O-RAN Alliance. O-RAN high-level architecture graphic. *O-RAN Minimum Viable Plan and Acceleration towards Commercialization.* June 29, 2021.

6. National Institute of Standards and Technology. Secure Software Development Framework. Accessed June 2022.

7. SAFECode. *Fundamental Practices for Secure Software Development, Third Edition*. October 28, 2019.

8. Microsoft. *Project Cerberus*. June 24, 2021.

9. Microsoft. *Comments on the U.S. Department of Commerce's Interim Final Rule to implement EO 13873, Securing the Information and Communications Technology and Services Supply Chain.* Filed Mar. 22, 2021.

10. National Institute of Standards and Technology. *SP 800-161 Supply Chain Risk Management Practices for Federal Information Systems and Organizations*. April 2015.

11. National Institute of Standards and Technology, National Cybersecurity Center of Excellence. *Supply Chain Assurance*. Accessed June 2022.

12. Microsoft. How it works: Azure AD Multi-Factor Authentication. February 7, 2022.

13. O-RAN Alliance. O-RAN high-level architecture graphic. *O-RAN Minimum Viable Plan and Acceleration towards Commercialization.* June 29, 2021.

Microsoft Azure

whitepaper

Bringing Cloud Security to the Open RAN