Juniper Networks | Driven by Experience™

White Paper

KEY CONSIDERATIONS FOR ASSURING DIFFERENTIATED, END-TO-END 5G SERVICES

Automated active assurance ensures that operators can turn up new 5G network slices properly and meet committed quality levels

TABLE OF CONTENTS

EXECUTIVE SUMMARY

5G is not only a technology; it provides fundamentally new capabilities compared to previous generations of mobile networks. For network operators, 5G provides a means for differentiated and guaranteed services. Operators can dynamically deliver quality-assured, diverse services over a common, shared infrastructure, all the way from the radio across transport networks to local and global data centers. Differentiated and guaranteed services make it possible to define and deliver committed service-level agreements (SLAs) for use cases targeted to specific industry verticals, such as media/entertainment, automotive, public transportation, e-health, and energy/utilities. 3GPP Release 16[1] specifies four standardized slices for enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), Massive Internet of Things (MIoT), and Vehicle to Everything (V2X).

This white paper primarily addresses the challenges that need to be addressed to win in the 5G race, and it details critical assurance requirements for managing B2B2x SLAs and guaranteeing service and slice quality. Based on the characteristics of a 5G network, test, assurance, and SLA management must be a vital part of provisioning and automation. It cannot be handled manually as an afterthought, as in classical infrastructure-centric and non-real-time assurance systems. Service-centric assurance, which focuses on assuring quality and reliability of services that are traversing the network[2], is also vitally important.

Introduction–A Technology and Business Model Shift with 5G

Besides a new air interface, 5G introduces several new technologies and concepts. The more important ones shown in Figure 1 are:

- Network slicing of user plane and control plane traffic to deliver differentiated services
- Disaggregated radio access network (RAN), with functions separated into centralized and distributed components
- Edge computing to provide for low-latency applications and to implement cloud RAN
- Virtualization of network functions and utilization of container technologies

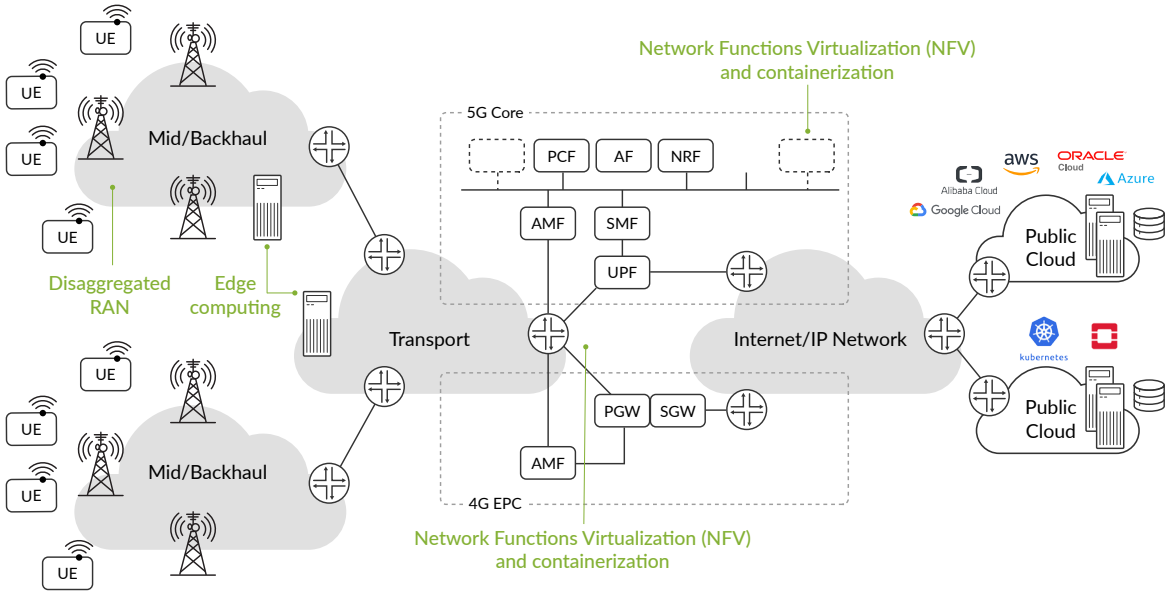

Figure 1: Overview of 5G and 4G radio access networks

[1] https://www.3gpp.org/release-16
[2] https://www.5gamericas.org/wp-content/uploads/2019/11/Management-Orchestration-and-Automation_clean.pdf

When armed with these new technologies, network operators can improve resource utilization, provide a greater degree of service flexibility, and improve time to market through automation. To ensure that 5G delivers the expected end-user quality of experience, network operators need to tackle new service assurance challenges introduced by these new technologies.

Challenges to Overcome to Become a Winner in the 5G Race

SLA agreements for individual slices and verticals put requirements on different key performance indicators (KPIs). For instance, an industrial application that controls manufacturing robots will require less bandwidth (but higher availability) compared to an augmented reality application that overlays real-time video of ancient buildings when sightseeing in Rome. Similarly, a remote surgery application with haptic feedback expects single-digit millisecond latency, compared to a taxi fleet management application that makes do with packets just coming through the network.

To be a player, and a winner in the 5G domain, three critical technical challenges must be addressed:

- **Dynamic provisioning of services and slices**. The network is a not a static pipe; rather, 5G service providers must support dynamic and fully automated slice creation.
- **Guaranteeing end-to-end service and slice quality**. The essence of a slice is the delivered real-time SLAs: the slice must be guaranteed at delivery and assured constantly. Customers will not accept bills that do not correspond to what is actually delivered.
- **Quickly isolating problems within network services' delivery chain**. When quality is compromised, quickly identify where the performance degradation comes from across the delivery chain of the end-to-end service so that the problem can be remediated before customers are impacted.

Ideally, active assurance enables a shift to a more effective service-centric operations model that is needed to meet service-level objectives.

New Demands to Assure Differentiated, End-to-End Services

Multiple new concepts in 5G introduce challenges for service assurance compared to the more static, non-virtual environment in traditional 4G networks. Table 1 details these challenges.

Table 1: New Technology Areas and Assurance Challenges

New Technology Area	Assurance Challenge
Network slicing	Slice instances are dynamically triggered and created through automated service provisioning. Service testing, quality-of-service (QoS) measurements, and operations handover can no longer be separate processes, sequentially scheduled and executed. Instead, they need full integration and instant execution. Network slice and network service monitoring must be automatically instrumented.
Edge computing	Data plane performance across the edge compute node depends on many factors such as the choice of network interface being used (virtio, SR-IOV, or PCI pass-through) as well as the behavior of "unfriendly" virtualized network function (VNF) workloads on the same edge node that impact the data plane performance, known as noisy neighbors.
Network Functions Virtualization (NFV) and containerization	Assuring data plane performance across VNFs and Cloud-Native Network Functions (CNFs) is more demanding than when custom-built hardware is being used. The dynamic nature of 5G makes it important to continuously validate performance and service quality, every time it is deployed or upgraded as part of life-cycle management.
Disaggregated RAN	To ensure low-latency services, it is critical to assure the performance of the IP/Ethernet network constituting the backhaul and midhaul networks[3, 4] connecting to the distributed baseband processing units (see Figure 2).

[3] https://www.ngmn.org/wp-content/uploads/Publications/2019/190412_NGMN_RANFSX_D2a_v1.0.pdf
[4] https://www.mef.net/service-standards/underlay-services/5g/

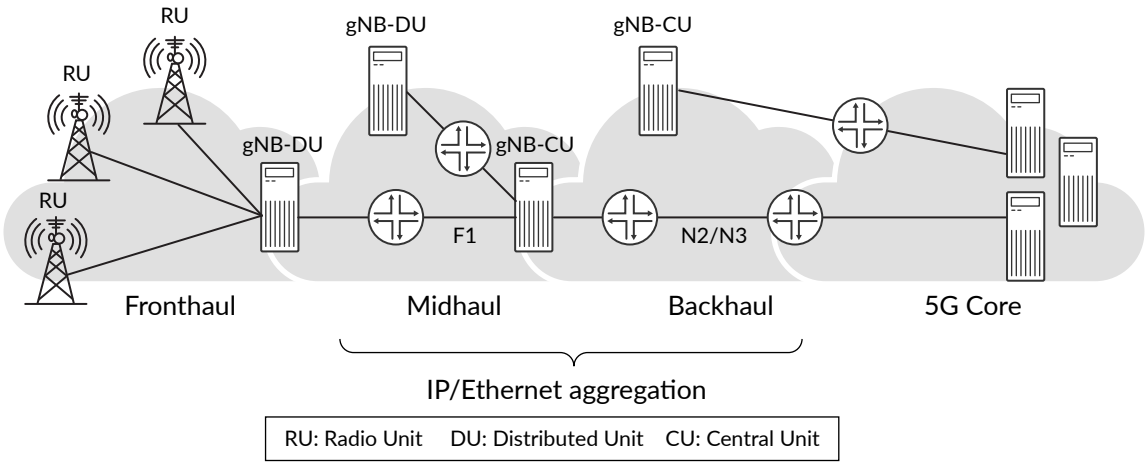

Figure 2: 5G end-to-end network services

Why Classical Assurance Does Not Meet All Needs for 5G

Classical telecom assurance solutions primarily focus on collecting information from the infrastructure, devices, and more recently the resource utilization of VNFs. Most solutions have been based on complex approaches to inferring the quality of the services based on what can be observed from devices. The reality is that device-centric counters and alarms correlate poorly with customer satisfaction. As illustrated in Figure 3, this causes a disconnect between what the customer experiences and what is seen by the network operations teams and service desks.

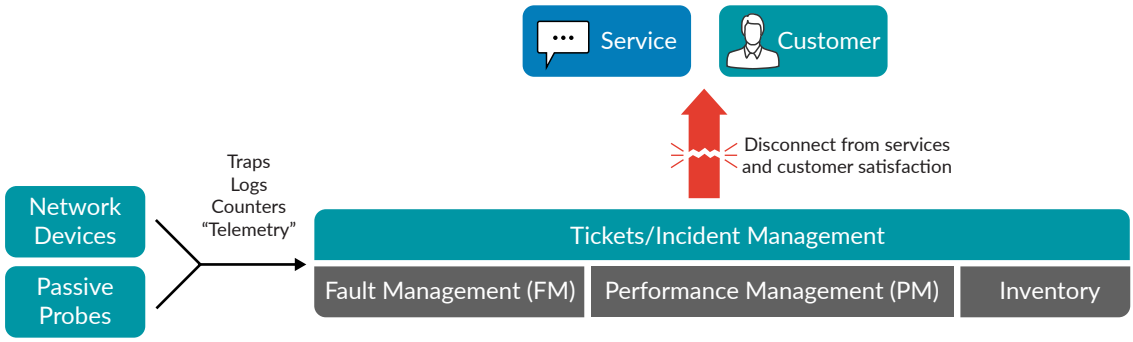

Figure 3: Traditional service assurance in telecom

In 5G, services and slices will be instantiated on demand in a highly dynamic environment. This presents many challenges for traditional assurance solutions that lack capabilities to provide a real-time view of the actual end-to-end service quality in this fast-changing environment. For instance, it is not possible for traditional fault and performance management systems to answer questions like:

- Is the one-way delay of my URLLC service meeting the requirements?
- Can my massive IoT service access the applications in all public clouds?
- Is the TCP throughput for my eMBB service achievable as sold and announced?
- Can an enterprise slice deliver video conferencing and data bulk transfers simultaneously?

Incomplete and inadequate service validation creates frustration for subscribers, leading to damaged reputations and churn. Assurance solutions must change the focal point from infrastructure to actively measuring actual service quality.

Service-Centric Operations with Active Assurance

Active assurance enables a shift to a more effective service-centric operations model that is needed to meet service-level objectives. By shifting to Service-Centric Operations with Active Assurance, mobile operators can:

- Proactively measure and assure end-to-end service quality through the data plane to confirm that network slices support business objectives
- Understand customer experience from an end-user perspective using synthetic L2 to L7 traffic
- Shorten time to resolve problems
- Locate performance issues before customers are onboarded and impacted

Note: More on service-centric operations in the 5G era is captured on our resources page, including a related white paper, TM Forum webinar, and further reading materials.

An active assurance solution uses IP hosts, often called active test agents, located in the network to send and receive synthetic traffic. This makes it possible to actively confirm that network services work when configured and continue to work during their lifetimes. This brings data plane visibility of differentiated services in virtual and physical environments across all network layers.

The best way to actively assure mobile networks pre-5G has been, and still is, to actively monitor the mobile backhaul network, typically using a standardized reflection technology such as Two-Way Active Measurement Protocol (TWAMP). Most eNodeBs support this standard today, making it possible to assure a large mobile network from only a couple of test agents located in central data centers. The central test agents are sending TWAMP UDP packets toward eNodeBs, which reflect those UDP flows back to the test agents. This makes it possible to measure latency, jitter, and packet loss continuously, and in a granular way detect KPI threshold violations. As an example, Orange Egypt deployed an active assurance solution from Juniper for its backhaul network in record-breaking time over a single weekend.[5]

In 5G, this solution is somewhat limited since it does not measure and assure the users' data plane through the dynamic environment of the 5G network such as the software-based User Plane Functions (UPFs).

For active assurance deployment to be feasible in large-scale, automated 5G networks, it is preferred that the overall solution provides the capabilities outlined in Table 2.

Table 2: Active Assurance Deployment Capabilities and Requirements

Capability	Requirements
Centralized test and API monitoring	Network automation frameworks and orchestrators should have access to a central API to utilize active assurance capabilities across distributed active test agents.
Coverage of the full operational life cycle	To avoid complex integrations of multiple-point solutions, an active assurance solution should combine turn-up testing, ongoing real-time active monitoring, and troubleshooting into a single solution.
Zero-touch dynamic deployment	Active test agents, either containers or virtual machines (VMs), should be instantiated as part of service or slice creation. This provisioning should be fully automated and zero touch.
Small footprint and minimal resource requirements	Specifically at edge locations, there is a limited amount of compute and storage, which means that an active assurance solution must only allocate a fraction of available resources. Typically, this involves consuming only a single vCPU and executing in a few hundred MBs of RAM.
Measurement through the 5G data plane	The test agents must be able to send traffic through the 5G data plane. This means that the traffic must be encapsulated in the GPRS tunneling protocol (GTP) tunnel and traverse the network slice the same way as mobile phone (UE) traffic does.
Service chains compatibility	The test agent must support flexible networking so that it can act as a small VNF in the service chain. In this way, it gets full visibility into the data plane traversing individual VNFs in the service chain, as well as the complete service chain data plane KPIs.
Multilayer, L2–L7	To isolate issues with different protocol layers of the data plane, the solution needs to be able to mix, concurrently and arbitrarily, active traffic from the link layer (L2) to the application layer (L7).

[5] https://www.juniper.net/us/en/customers/orange-egypt-case-study.html

Capability	Requirements
Performance at scale	The solution should handle deployment of thousands of active test agents wherever there is compute available. Any active test agents should be able to scale to thousands of concurrent parallel streams or sessions to support use cases based on reflection technologies (TWAMP, Y.1731, UDP Echo, ICMP Echo) towards existing network elements.
Accurate timestamping and high resolution	To confirm one-way delays in midhaul networks, measuring with sub-millisecond accuracy and precision is a requirement. The solution must be able to use hardware timestamping on physical network interfaces.
IPv6-only support	Many modern networks are deployed without IPv4, which means that the active assurance solution must support environments where only IPv6 is available.

Juniper® Paragon™ Active Assurance meets all the critical capabilities and requirements highlighted in Table 2.

Active Assurance for 5G

There are three relevant use cases for active assurance in 5G environments:

- **Network slicing**—Confirm overall slice performance from user equipment (UE) to services located in public, private clouds, or edge clouds.
- **Midhaul/Backhaul**—Confirm network performance of the network interconnecting the RAN and the core.
- **Service-Based Architecture (SBA) network**—Confirm network performance and service availability of the 5G control plane functions.

For network slicing, 5G operators must be able to send traffic through the 5G user plane. In Paragon Active Assurance, this is possible by emulating the UE and gNB as part of the test agent.

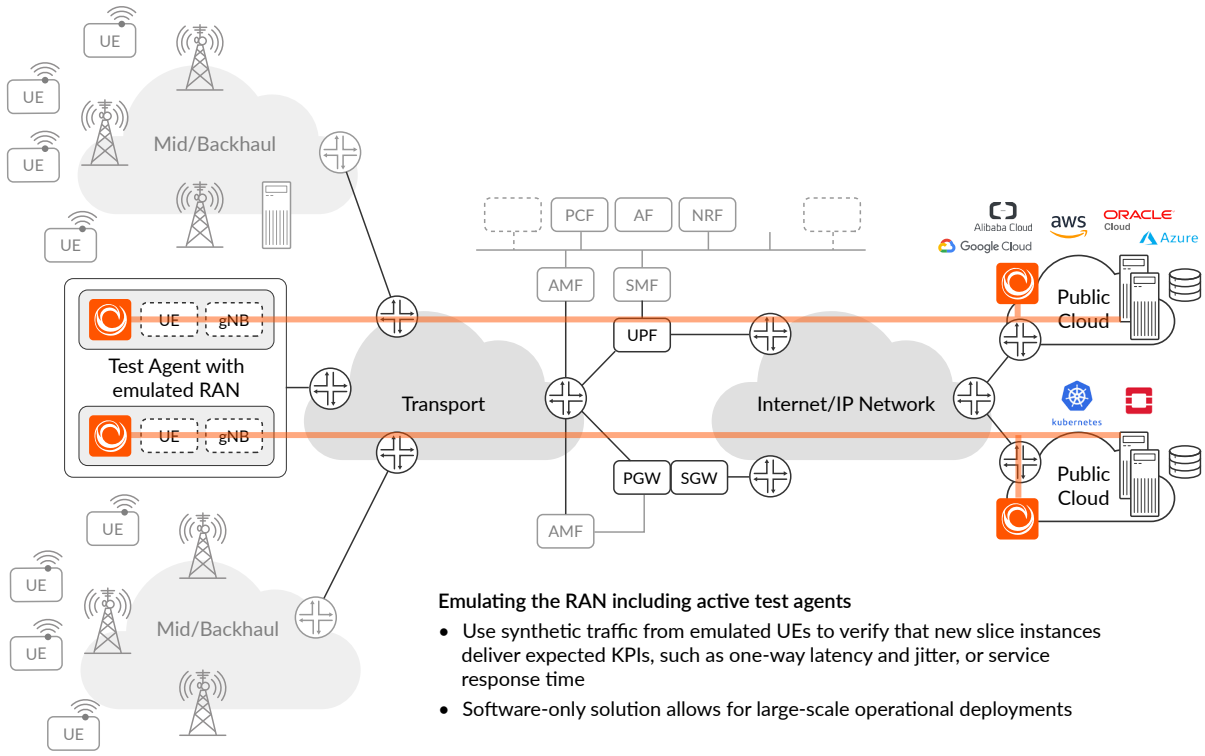

Figure 4: Active test agent combined with emulated RAN, for complete end-to-end KPIs across individual slices

This setup allows for the following important use cases:

- Ensures slices meet expected SLAs on the user plane, end-to-end from UE to apps
- Confirms successful instantiation of new differentiated slices
- Uses active traffic across all involved UPFs to discover performance degradations before customers notice

To assure the midhaul/backhaul, depending on the available compute locations, testing of services via the user plane can also be combined with using reflection technologies towards the RAN devices (as is done today in 4G networks). See Figure 5.

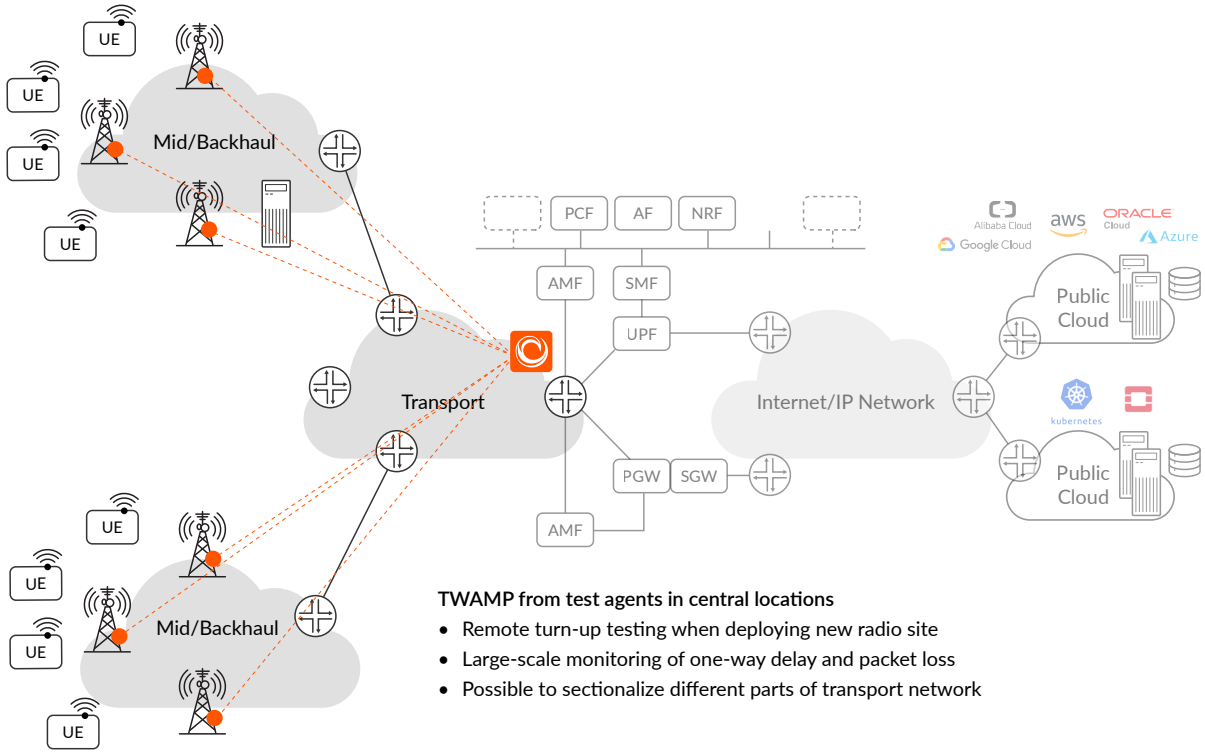

Figure 5: A single active test agent covering a region interconnecting the RAN with the core

The test agent will, in addition to emulating the UE/gNB and sending traffic via the network slice over the N3 interface via the UPF out to the Data Network (DN) to remote public clouds on the Internet, also be able to complement with Y.1731 or TWAMP measurements for validating the midhaul/backhaul network in case no compute is available further out in the network.

The further out in the network the test agent can be deployed, the better. The most natural locations are at the RAN site where the gNB is located, and if that site cannot run containers, the second best is at the distributed unit (DU) sites. Otherwise as a last resort, the centralized unit (CU) can use TWAMP out towards the gNB. Note that the wireless radio part is not covered by dynamic active assurance.

Active assurance is also valuable in the mobile core SBA. As illustrated in Figure 6, active test agents are deployed as containers in the Kubernetes cluster implementing the SBA control plane. Typically, test agent containers will run as Kubernetes sidecars as CNFs to support the following use cases:

- Confirm Kubernetes cluster networking in an SBA
- Validate network performance for individual control plane slices
- Use synthetic HTTP requests on the control plane to monitor SBA network functions and provide alerts if they suddenly fail

The SBA might be distributed across multiple locations, and while it is easily forgotten, a well-functioning network connecting all the mobile core workloads is of highest importance.

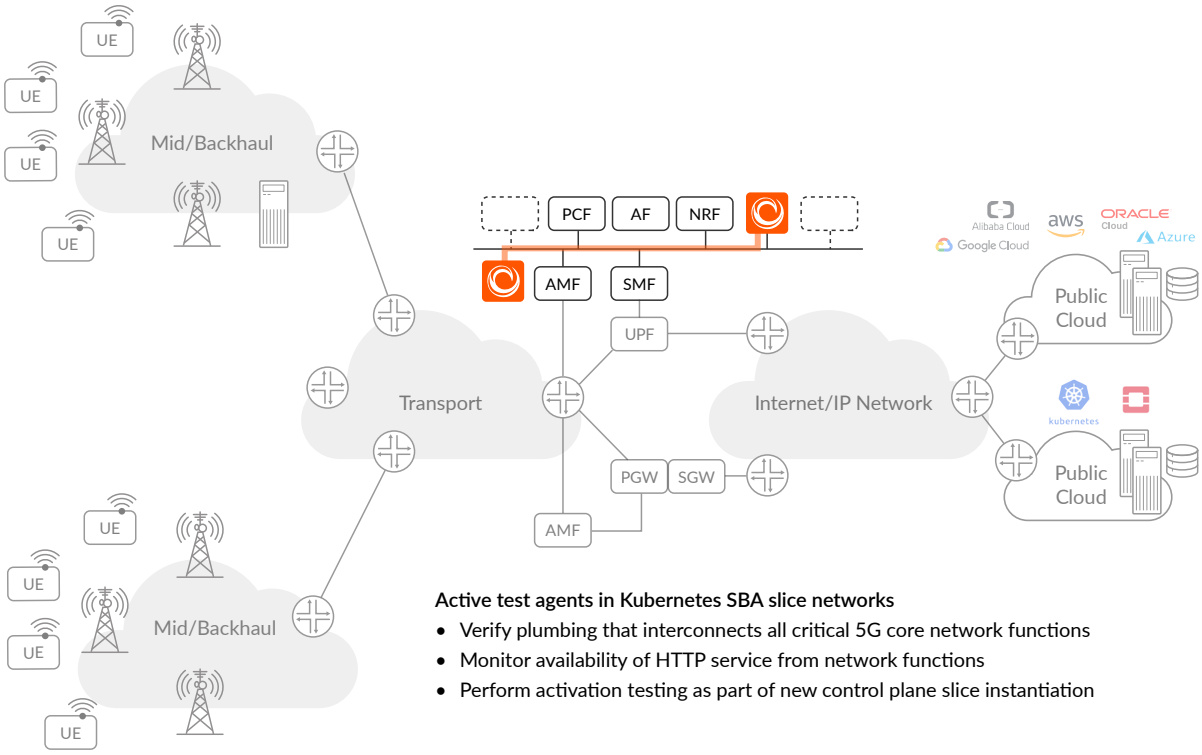

Figure 6: Active test agents deployed as part of a Kubernetes cluster for the 5G control plane

Many times, 5G deployments involve both VM and container environments. This means that an active assurance solution needs to provide test agent capabilities for both compute technologies. As illustrated in Figure 7, the edge node might run the virtualized Distributed Unit (vDU) as a VM for performance reasons. In this scenario, it would be desired to run a VM-based active test agent alongside the vDU. Similarly, a regional data center might only run containers on Kubernetes, while an active test agent needs to run as a sidecar deployment. It is important that test agents are interoperable and compatible regardless of VM or container deployment type.

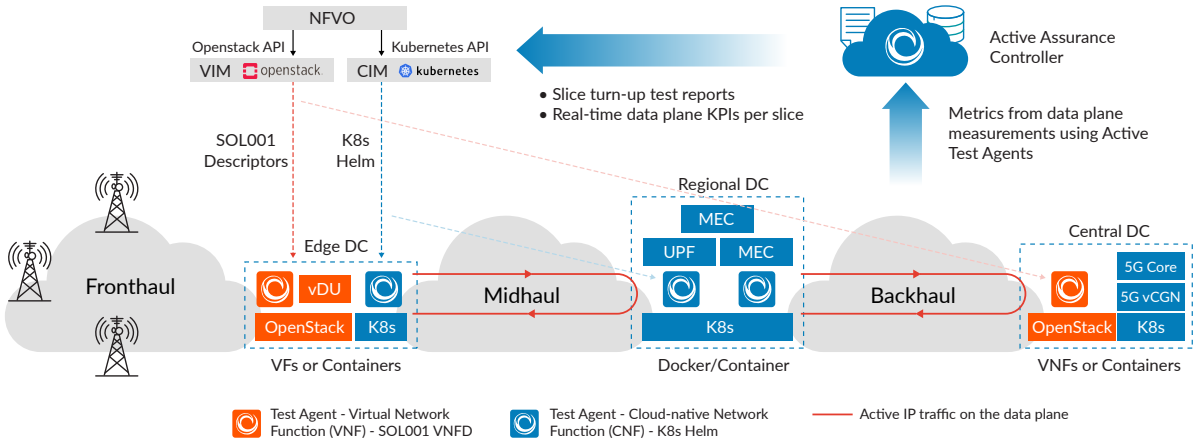

Figure 7: Example of a mixed environment with both OpenStack and Kubernetes compute environments

Automated Active Assurance Workflow

True 5G automation needs test and real-time slice/service monitoring capabilities built into the orchestration workflow. The orchestrator must be able to validate that the slice works at the data plane, not just that the individual resources are healthy. This guarantees that it is measuring what matters for the customer, good service, and it requires measurements on the data plane. It is important to measure KPIs for the data plane packets because that is what makes a slice work or not work.

And, the orchestrator needs to measure and monitor active and synthetic traffic. This ensures that the service can be tested at delivery before customers are onboarded. Also, ongoing SLA monitoring needs to be based on continuous, synthetic traffic. How else would it be possible to know, for example, that the latency in a mine is good enough for an autonomous truck to enter? Passive probes only reveal poor performance after the fact. Fault management and performance management systems do not even see the service or slice, as they only see the individual infrastructure devices, resources, and VNFs.

Figure 8 shows three loops starting and ending at the orchestrator:

- The first loop tells the orchestrator that committed service KPIs could not be obtained.
- In the second loop, the orchestrator attempts to correct broken configurations or non-optimal configurations.
- In the third loop, the orchestrator uses conclusions from correlated root-cause analytics to attempt to restore the service using non-faulty or non-congested resources.

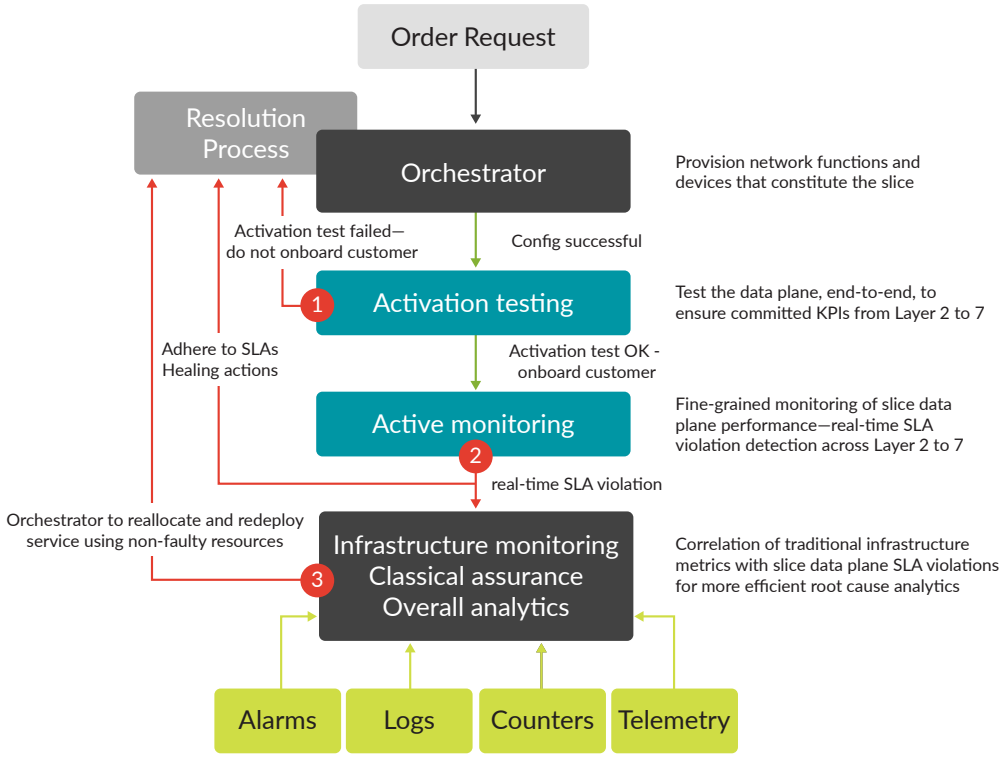

Figure 8: Logical orchestration workflow across orchestrator, active testing/monitoring, and classical assurance

As shown in Figure 9, active traffic on the data plane bridges the gap between classical assurance and network orchestration. When the active solution detects an issue in real time, classical assurance systems (to the left) can analyze an underlying fault, while the orchestrator (to the right) can investigate if it is a configuration issue. Note that a common cause of network performance degradations is related to configuration mistakes, and not faults as such. Therefore, device-centric classical assurance systems do not help in detecting nor analyzing these kinds of problems.

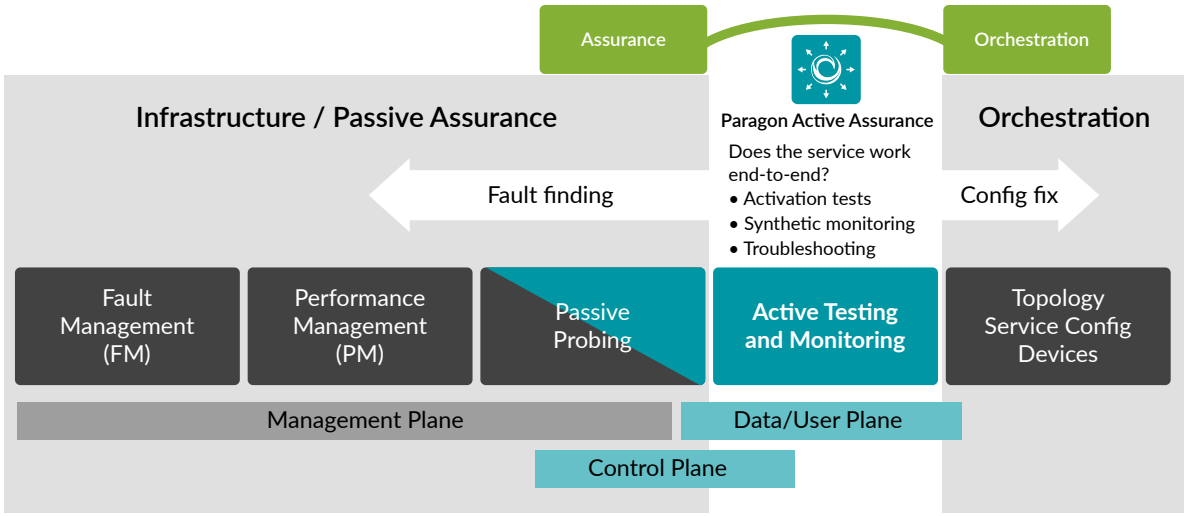

Figure 9: Automated active traffic on the data plane bridging the gap between classical assurance and orchestration

Conclusion

Extreme automation, combined with fully automated active assurance of individual slices, is front and center in any successful 5G network deployment.

Network slicing enables service differentiation and Network as a Service (NaaS) offerings supporting a diverse set of guaranteed SLAs and KPIs. Old, static, and infrastructure-oriented approaches for service assurance are doomed to fail in the ever-changing virtual environment that forms the foundation for today's 5G architecture.

For network operators to have a profitable 5G business, they need the capability to ensure that new network slices are properly turned up and that individual slices meet committed quality levels during the slice's lifetime. To offer profitable services, a new active approach for service assurance is needed. The new approach is characterized by:

- The ability to orchestrate the assurance solution directly into the service fabric
- Testing services as part of delivery
- Exposure of a centralized API for service orchestrators to consume for service activation testing, ongoing real-time data plane monitoring, and remote troubleshooting
- Real-time feedback to orchestrators and analytics frameworks to achieve closed-loop orchestration

Taking an active approach to assure your differentiated service offerings is an important step in becoming a winner in the 5G race.

About Juniper Networks

At Juniper Networks, we are dedicated to dramatically simplifying network operations and driving superior experiences for end users. Our solutions deliver industry-leading insight, automation, security and AI to drive real business results. We believe that powering connections will bring us closer together while empowering us all to solve the world's greatest challenges of well-being, sustainability and equality.

Corporate and Sales Headquarters
Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER (888.586.4737)
or +1.408.745.2000
www.juniper.net

APAC and EMEA Headquarters
Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk
Amsterdam, The Netherlands
Phone: +31.207.125.700