Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING
WHITE PAPER

5G Network Visibility:
Why CSPs Need Visibility across All Domains, Technologies, and Vendors

A Heavy Reading white paper produced for NETSCOUT

NETSCOUT®

AUTHOR: JAMES CRAWSHAW, SENIOR ANALYST, HEAVY READING

INTRODUCTION

In the world of network engineering and operations, network visibility is a much sought-after yet hard-to-define goal. Generally speaking, communications service providers (CSPs) want more of it, provided that the monitoring activity itself does not put an excessive load on the network or become cost-prohibitive due to the number of probes or the amount of storage required.

With the introduction of 5G, operators are looking for greater levels of visibility to help them manage increased complexity, new interfaces, and the separation of control and user plane (CUPS). Operators' experience with 3G and 4G network technology will only get them so far; 5G introduces new challenges.

In the initial non-standalone (NSA) mode of 5G deployment, New Radio (NR) cells are combined with 4G/LTE radio cells to provide dual connectivity with a common 4G core network (Evolved Packet Core). Here, the control and user plane can be separated between LTE and NR to minimize the number of handovers between cells. In practice, when adding small 5G cells within a larger 4G macro cell, the 4G cell handles the control plane traffic.

User equipment, however, always connects to whichever radio delivers the best user plane connection: 5G when inside the small cell or 4G when outside the small cell's range. The dual-mode nature of NSA means that frequency, power, and antennae orientation must be configured for both 4G and 5G to ensure that overall coverage and performance is optimized. To support these early NSA scenarios, operators need visibility into subscriber session performance across 4G core and 5G radio.

In time, a new 5G core will be deployed and connected to the 5G radio, leading to a longer-term standalone (SA) solution. In the SA model, the 5G NR radio cells will handle both the control plane and user plane. The SA option should simplify management by allowing an independent network using normal, inter-generational handover between 4G and 5G.

To support this 5G evolution path, operators need analytics capabilities that provide insights across the full life cycle from design to rollout to ongoing operations. NSA to SA is not the only transition operators must anticipate. Software releases by network equipment suppliers will come at a faster cadence than with previous technology generations, and each one will need testing to ensure network efficiency and customer experience remain high.

This Heavy Reading white paper discusses the importance of visibility into end-to-end subscriber sessions that span multiple network domains, including physical and virtualized (virtual machine [VM]- and container-based), and multiple vendors. It goes on to describe how visibility should be turned into smart data for higher level systems and tangible benefits. Heavy Reading discusses the use cases of smart data-based analytics across the CSP organization and throughout the life cycle of 5G deployments. The paper also examines the detailed 5G data analytics functions that form part of the evolving 5G Public Private Partnership (PPP) architecture and the importance of cloud-native design for network monitoring and analytics.

NEED FOR MULTI-VENDOR, DOMAIN, AND TECHNOLOGY VISIBILITY

Today's networks have many borders; they have borders between domains (core, radio access network [RAN], transport, access, data center, etc.) and between technologies (3G, 4G, 5G, cloud, etc.). Each one of these infrastructure types comes with proprietary monitoring tools supplied by its vendor. Stitching these networks together to get a holistic view presents a significant challenge. The problem becomes exacerbated with the move to network functions virtualization (NFV – itself a prerequisite for 5G). Operators need to monitor the physical infrastructure (switches, routers, etc.) and the servers that support virtualized network functions in addition to the software stacks that support these (VMs, virtual infrastructure managers, orchestrators, etc.).

The virtual network functions (VNFs) themselves (IMS, EPC, SBC, etc.) also require vendor-agnostic, lightweight monitoring that can provide the context of the service chain and subscriber session. Without visibility across all these layers, CSPs lack the insight needed to optimize performance, get end-to-end visibility of services (voice over IP [VoIP], Internet of Things [IoT], network slices, etc.), and deliver a premium customer experience.

For many CSPs, existing analytics solutions cannot integrate the diversity of data sources. As a result, they often end up with huge datasets that take a long time to assemble and analyze, precluding the real-time analytics required for closed-loop automation. Few of today's systems enable the operator to understand the quality of experience at the application layer as opposed to the network layer.

> *"If networks are monitored in real time, closed-loop control becomes possible, delivering automated self-healing and elasticity."*
>
> *– Jose Domingos, OSS Chief Assurance Architect, BT, in Enterprise Networking magazine, April 2019*

Better Visibility Required to Support NFV

5G is being implemented in tandem with the move to NFV. As such, it will require the monitoring and correlation of performance data across multiple layers like NFV infrastructure, virtual infrastructure managers (OpenStack, Kubernetes, etc.), VNF managers, and the VNF applications themselves. Monitoring tools for NFV will need to collect data from new sources such as servers, hypervisors, containers, and a plethora of VNF vendor-specific application programming interfaces (APIs).

NFV poses several new service assurance challenges:

- The operating environment will be highly dynamic partly due to the inherent nature of NFV and partly due to the increased service agility it enables and the concomitant variation in customer requirements (e.g., bandwidth on demand).

- Problems in the physical infrastructure layer (NFVI) will lead to problems in the virtualized layer (VNFs), requiring much closer cooperation between the IT and network operations teams.
- The increased level of automation that NFV promises will require more reliable monitoring systems to provide the necessary closed-loop assurance.

CSPs must have visibility down to the VNF or container bin with virtual instrumentation that also provides the context of the service chain and subscriber session. This virtual instrumentation must be lightweight. The compute resources must be reserved and prioritized to deliver network quality of service (QoS) such as Ultra-Reliable Low-Latency Communication (URLLC) and Enhanced Mobile Broadband (eMBB) and/or support high device numbers (Massive Machine-Type Communications [mMTC]).

Increased Visibility Required to Support Network Slicing

According to a recent Heavy Reading survey, almost all CSPs are considering network slicing. Roughly half plan to launch it within 3 years of commercial 5G launch and around a fifth plan to do so within 1 year. However, while many operators are talking about network slicing, very few have figured out how to operationalize it. One of the greatest challenges they face is lack of a strong monitoring capability for network slicing. To guarantee service-level agreements (SLAs) for network slicing, they will need to monitor traffic characteristics and performance (e.g., data rate, packet drop, and latency) and end users' geographical distribution and perform per session/user/slice instance-based monitoring.

SMART DATA TO FEED MULTIPLE ANALYTICS SYSTEMS

There are multiple sources of data within networks today. In fact, many operations teams complain of data overload. They have an abundance of device statistics (traffic counters, CPU/memory, usage, etc.), log data, flow data (NetFlow, sFlow, Jflow, etc.), session records (XDRs showing video metrics, voice metrics, etc.), and packet trace/capture records (e.g., deep packet inspection [DPI]). Collecting all of this data can rapidly lead to an operator's data lake turning into a data swamp.

> *"The best strategy for data lakes is to only collect data that is useful now. Data loses its value over time and if you can't find what you're looking for in the mess that is the data swamp, it's pointless to keep adding to it. Projects should only go after sources that can provide useful solutions to clearly defined business problems."*
>
> *– Nick Ismail, Information Age, June 2017*

Instead of creating data swamps, operators should turn raw data into more actionable "smart" metadata that is contextual (who, what, where, when), relevant, structured, compact, timely, and of requisite granularity. Such intelligent filtering might reduce the volume of networking monitoring traffic by as much as a factor of 100, leading to savings in network utilization, as well as server and storage resources.

Real-time, service-contextual, smart metadata can serve multiple groups within the CSP:

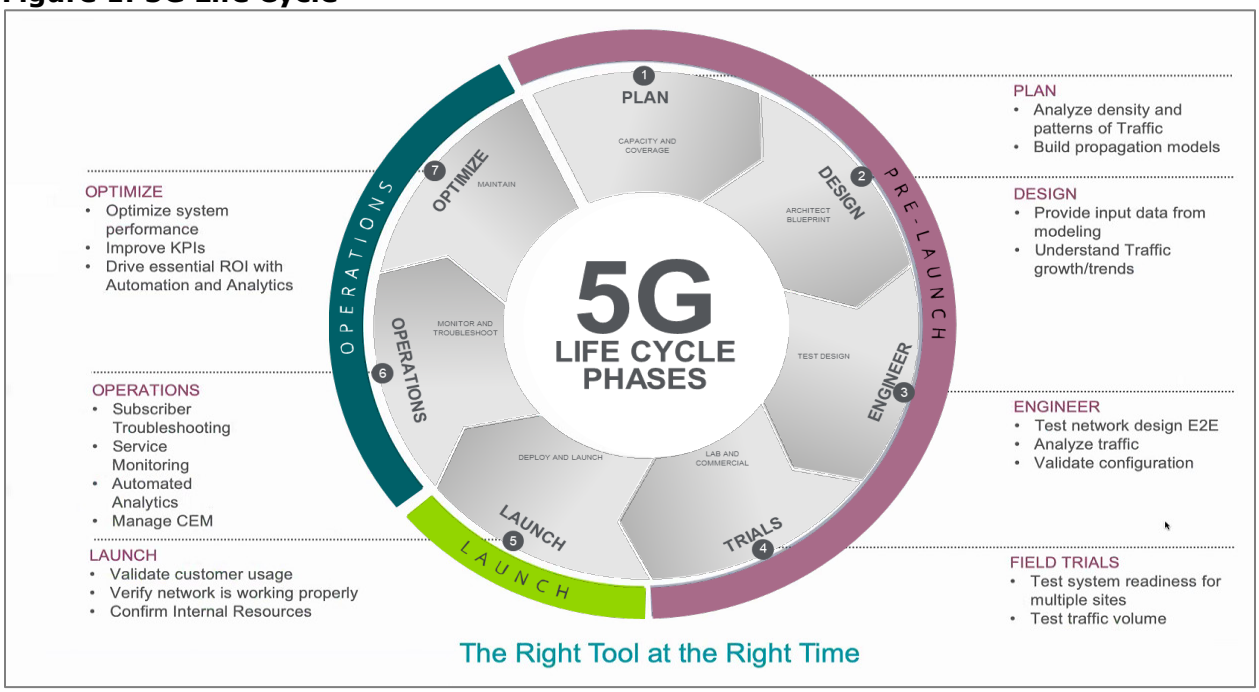
- Network operations/service assurance
- Customer care
- Marketing
- Network planning
- Security operations
- Other groups such as service delivery, business operations, targeted advertising, etc.

When deploying 5G, CSPs need analytics to understand what is going on in the network and have insight into the performance of any given service (eMBB, URLLC, mMTC, etc.). Analytics can also help decide which markets to roll out 5G in and when. For example, the operator would check the up/downlink throughput, Transmission Control Protocol (TCP) latency, and attach rate in the network before launching 5G in a particular region. Analytics could help identify root causes for any 5G vendor interoperability issues, including handsets. Providing network monitoring data to security systems could give operators a new tool in their armory.

ANALYTICS ACROSS THE 5G LIFE CYCLE

Another way of thinking about the analytics of applications for 5G is to look at the life cycle for a new network technology, as shown in **Figure 1**. The initial phase is planning; here we need to analyze the density and patterns of existing network traffic and build propagation models. In the design phase, we need analytics to understand network traffic growth trends. At the engineering stage, we need analytics to test the network design end-to-end, analyze traffic, and validate configurations. Once operators reach field trials, they need analytics systems to test system readiness. At the launch phase, analytics is needed to validate customer usage and verify that the network is working properly. In the ongoing operations phase, operators need to conduct service monitoring, manage the customer experience, and troubleshoot issues. There is an ongoing optimization effort that will require analytics to improve key performance and quality indicators (KPIs and KQIs) and overall system performance. The cycle will continue with each new release of the 5G standards and compliant vendor software.

Figure 1: 5G Life Cycle



Source: NETSCOUT

The analytics solution should not just tell the operator what is going on in the network. It should also provide diagnostics that indicate why it happened, make predictions about how a KPI or KQI will behave in the future, and make recommendations about what actions should be taken to maximize network performance and customer experience.

5G DATA ANALYTICS FUNCTIONS

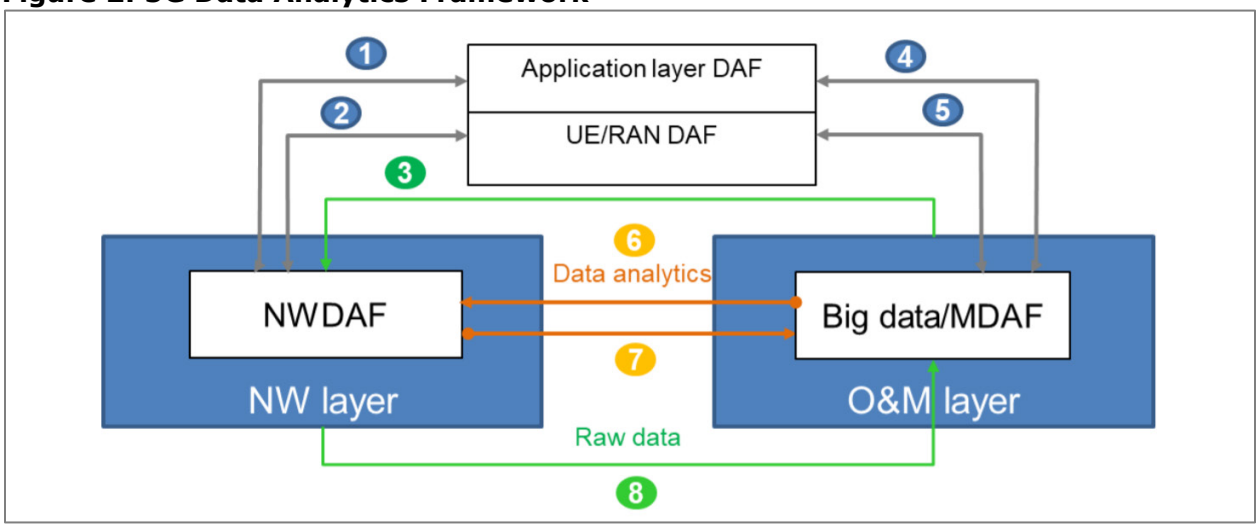
According to 5G PPP Architecture Working Group View on 5G Architecture Version 3.0, the analytics framework for the 5G core network architecture considers data analytics capability at various layers and introduces data analytics functions (DAFs) into the following:

- Core network domain (network data analytics functionality [NWDAF])
- Big Data and management and orchestration (Big Data/MDAF)
- Application function level (AFDAF)
- User equipment/RAN-DAF
- Data network (DN-DAF)

An example use case of one of these DAFs would be to send KPIs to a policy control function to decide about security on a noisy node or determine if QoS is unequal across subscribers in a cell.

Each logical data analytics module is implemented as multiple instances for different use cases and purposes. For instance, the Big Data module in the management and orchestration layer could be implemented as multiple instances per domain (e.g., RAN data analytics, VNF data analytics, etc.) at different levels (e.g., cross/intra domain). This framework allows for a dedicated data analytic module design at different layers, also enabling cross-layer optimization. **Figure 2** depicts the overall integrated data analytics framework.

Figure 2: 5G Data Analytics Framework



Source: 5GPP

The data analytics parameters are classified as follows:

- **User equipment/session-related:** E.g., prediction of user location to optimize handover
- **Network-related:** E.g., average channel quality in the RAN

- **Service-related:** E.g., vehicle-to-X (V2X) application layer analytics
- **Management-related:** E.g., performance management (PM) and fault management (FM) analytics as introduced in 3GPP SA5
- **Cloud-related:** E.g., the load and availability of computational resources, which may affect the decision about location of VNFs on cloud platforms

The granularity of data collection in the time domain is a major factor in the amount of analytics data that must be transported, processed, and stored. This granularity can be categorized as follows:

- **Real-time:** For real-time operations such as channel prediction in millisecond time scale. This is challenging, as additional processing might be required and the overhead may affect performance.
- **Near real-time/non-real-time:** The analytics is performed every few seconds, minutes, or even hours. This may apply to certain predictions such as load distribution in a geographical area. In Open-RAN (O-RAN), near real-time operations have been defined to capture operations like QoS management, traffic steering, and mobility management.
- **On demand:** This can apply to both real-time and non-real-time analytics and is the case when the operator requires analytics as a service for a given area or time window to meet the requirements of a network slice.

The types of analytics can be categorized as follows:

- **Descriptive:** Explains what is happening presently based on current or live data.
- **Diagnostic:** Examines past performance to determine what happened and why.
- **Predictive:** Analyzes likely scenarios of what might happen in the future.
- **Prescriptive:** Suggests what actions should be taken to resolve a diagnosed problem or avert a predicted issue.

THE IMPORTANCE OF CLOUD-NATIVE SOLUTIONS

When selecting an analytics solution, operators should ensure that it is architecturally as well as functionally sound. The state of the art for software design is cloud native. Cloud native describes software that is designed, developed, and optimized to exploit cloud technology (i.e., distributed processing and data stores). According to Ken Owens,[*] Cloud Native Computing Foundation (CNCF) Technical Oversight Committee representative, cloud-native applications are a combination of existing and new software development patterns. Existing patterns include software automation (infrastructure and systems), API integrations, and service-oriented architectures. New cloud-native patterns include microservices architecture, containerized services, and distributed management and orchestration.

The CNCF's own definition of cloud native describes it as such:

- Microservices-oriented (loosely coupled with dependencies explicitly described)
- Container packaged (high level of resource isolation; fosters code and component reuse; simplifies operations)
- Dynamically managed (central orchestrator improves resource utilization)

Microservices-Oriented

According to Matt Stine,[†] global chief technology officer (CTO) of Software Architecture at Pivotal, microservices represent the decomposition of monolithic business systems into independently deployable services that do one thing well. That one thing usually represents a business capability or the smallest, atomic unit of service that delivers business value. Microservice architectures enable speed, safety, and scale in several ways:

- Each microservice can be updated independently, enabling more frequent and rapid updates.
- Developers can work in parallel (adding more people to a tardy software project delays it further).
- Independent scaling of services (efficiency) is possible.

As Eyal Felstaine and Ofer Hermoni of Amdocs explain in their chapter of *Building the Network of the Future*, the benefit of cloud native is that it simplifies and automates provisioning, upgrading, scalability, and graceful shutdown of software. A key assumption is that cloud-native software is to be stateless. It relies on backing services for any aspect of long-term memory such as state and data. Stateless applications put less load on the server, but the tradeoff is that they require additional information in every request or additional communications to a separate, fast access database (e.g., NoSQL on solid-state memory).

---

[*] *DevNet Create*, "Developing Cloud Native Applications," Ken Owens, February 2017.

[†] Matt Stine, *Migrating to Cloud-Native Application Architectures* (Massachusetts:O'Reilly Safari, 2015).

Container Packaged and Dynamically Managed

Container images such as those prepared via Docker are the common unit of deployment for cloud-native application architectures. Such container images are then instantiated by scheduling solutions like Kubernetes. Public cloud providers like Amazon and Google also provide solutions for container scheduling and deployment. Containers leverage modern Linux kernel primitives such as control groups and namespaces to provide similar resource allocation and isolation features as those provided by VMs with much less overhead and much greater portability. U.K. financial newspaper *Financial Times* was reportedly able to reduce its Amazon Web Services (AWS) server costs by 80% while moving to a much more stable tech infrastructure by adopting containers.

Service Mesh – Enterprise Service Bus for Microservices

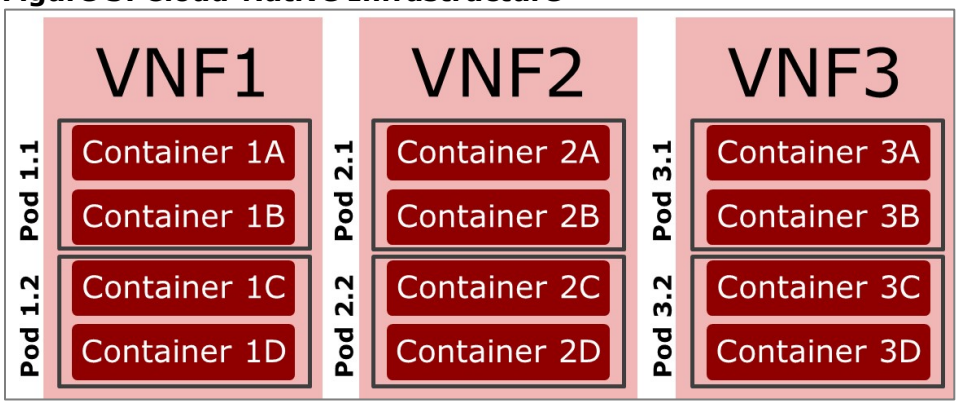
Interconnections between microservices can increase operational complexity (transaction processing, data locking, deadlock detection, critical races, and debugging). Inter-microservice communications increase processing load, communications load, and latency and create a programming bottleneck and a single point of failure. To address this problem, a layer seven networking concept called service mesh has been introduced. A service mesh (e.g., Istio) provides a uniform way to connect, secure, control, and observe microservices:

- **Connect:** Controls the flow of traffic and API calls between services
- **Secure:** Ensures authentication, authorization, and encryption of communications between services
- **Control:** Applies and enforces policies (e.g., resource distribution)
- **Observe:** Enables tracing, monitoring, and logging of services

Putting It All Together

Figure 3 shows some of the cloud-native infrastructure concepts as a whole. Each microservice runs in its own container. Multiple containers, plus any local storage (for state), are encapsulated into a pod. Each pod has its own IP address, and communications between pods take place using a service mesh such as Istio. Groups of pods run on nodes that could be bare metal or a VM. Each node contains the services necessary to run pods and is managed by Kubernetes master components.

Figure 3: Cloud-Native Infrastructure

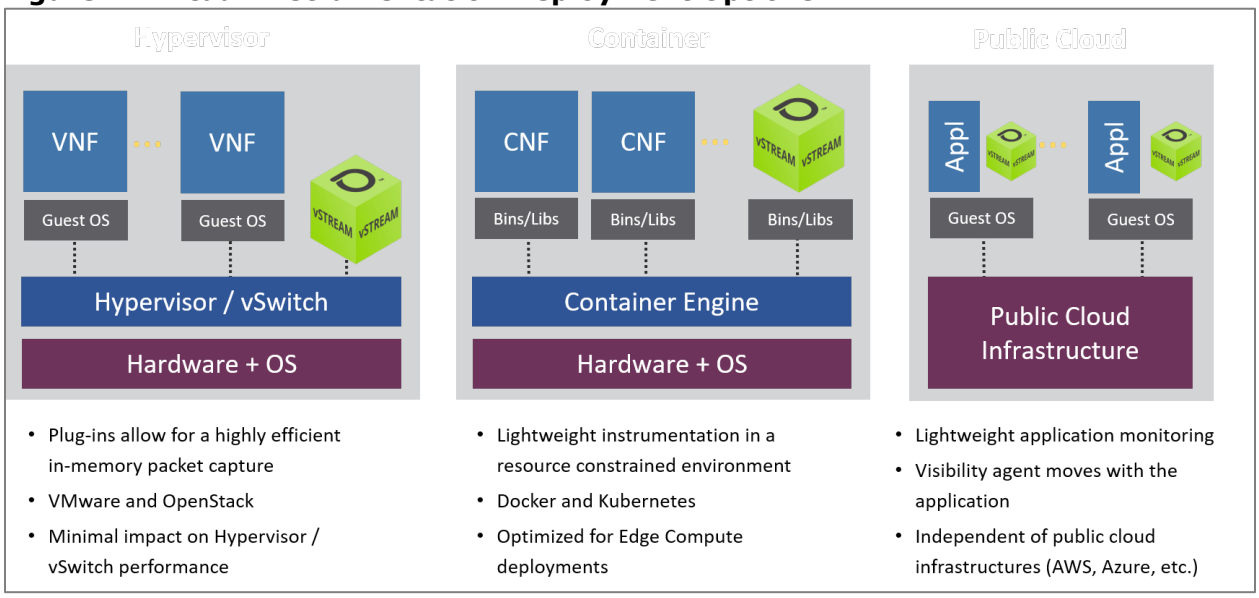


Source: Heavy Reading

How Cloud Native Benefits Network Monitoring

With a cloud-native approach to network monitoring software design, virtual probes can be deployed to run in containers within the public cloud or even within hypervisors, as shown in **Figure 4**. Software-based probes are generally cheaper than physical network interface cards. The move from VMs to containers should lead to an even lower total cost of ownership. However, care must be taken with virtual probe design to ensure it is not too compute resource-hungry, given that large numbers of probes may need to be deployed.

Figure 4: Virtual Instrumentation Deployment Options



Source: NETSCOUT

CONCLUSIONS

Industry expectations for 5G are high in terms of data speeds and latency, as well as new service enablement. To meet these expectations, network quality matters more than ever. While 3G network optimization typically focused on KPIs and 4G focused on KQIs of services such as VoLTE, 5G will shift the focus to onerous SLAs associated with network slicing and customer experience indicators (CEIs). These CEIs represent a consolidation of KPIs and KQIs to provide an overall "satisfaction" indicator that reflects the perception of a user or device and the service they receive. CEIs will be part of the automated analytics needed for 5G. To capitalize on 5G, CSPs need consistent, high fidelity, real-time visibility across all layers of the network, including both physical and virtualized domains. To support 5G, networks must be monitored as they are deployed and on an ongoing basis, not as an afterthought.

But 5G is not an island. Operators need visibility across 5G, 4G, Wi-Fi, and even fixed access networks. Monitoring should cover the RAN, core, data centers, and public cloud infrastructure. Operators need visibility into the NFV infrastructure, the virtual infrastructure manager, and the orchestration and automation systems (ONAP, OSM, etc.). Real-time network monitoring is the feedback that enables closed-loop automation.

Virtualization and containerization are still quite nascent for telecom networks; operations teams will need new tools to help them monitor these environments. This monitoring must cover the VNFs (IMS, EPC, SBC) in a service chain and even down to individual, end-to-end subscriber sessions.

Monitoring systems for 5G depend on smart data that is contextual, timely, and has a unified data model. Analysis of the data should be done close to collection; only high level digests should be passed on to clients such as an orchestrator, not the raw data itself. The analysis should enable operators to move from proactive to predictive monitoring and self-optimizing networks.

VISIBILITY WITHOUT BORDERS: 5G SERVICE ASSURANCE EXPERIENCE

This section was written by NETSCOUT.

Agility, speed and mobility – this is what your 5G network must deliver to enable the digital revolution, the Internet of Things, smart cities, autonomous vehicles, virtual reality, and other new and yet unthought of services. You need to transform your network to a virtualized infrastructure; move services from 3G and 4G to 5G; all while protecting your business against the ever-increasing risk of disruptions and cyber threats. You must respond to the rapidly changing needs of your customers, users, clients, and patients, and do it faster and better than your competition.

You have entered uncharted territory with the cloudification of the network in 5G bringing cloud RAN, network slices, and containers for the mobile edge. You will be introducing new millimeter wave and greatly expanded MIMO to the RAN. Your data center is evolving from traditional, physical network elements to distributed compute resources that must be orchestrated for automation. Together this multi-layered, myriad of technologies, products, and services must all work together seamlessly.

This fragmented network framework is causing a loss of end-to-end visibility and expanding your cyber-attack surface. To regain control and reduce these risks you need a radically new approach.

At NETSCOUT, we have made an unprecedented investment of time, money and resources to solve this challenge and bring to market a practical solution that produces a common view of your services across multi-generation, multi-domain technology and organizational boundaries, across all networks, all locations and all users.

At NETSCOUT, we have a practical solution we call *Visibility without Borders*, delivered by our unique *Smart Data* technology that produces a common view of your network across technology, organizational boundaries, all locations, and all users.

With NETSCOUT's carrier-grade 5G solution, carrier service providers gain visibility end-to-end, with software-based monitoring anytime, anywhere across physical, hybrid, and cloud environments.

As guardians of the connected world, our mission is to empower you to confidently accelerate your move to 5G and harness those benefits faster.

That's Visibility without Borders!