WHITEPAPER

Cloud-native 5G – Observability and Operations at Scale

How to capture and act on the huge volumes of performance data generated by a 5G core network

Hewlett Packard Enterprise

Published by

MOBILE WORLD LIVE

Executive summary

Employing thousands of microservices, cloud-native 5G networks generate massive volumes of valuable data. Mobile operators need to collect and analyse this data to check that their systems are performing as expected and immediately identify any issues. To do that, they need effective observability tools. Deep and accurate observability will ultimately help operators to automate 5G networks, and thereby improve both efficiency and performance.

Developed by Mobile World Live in partnership with Hewlett-Packard Enterprise, this paper explores why cloud-native 5G networks need to be observable, what challenges need to be overcome and what kinds of tools are involved.

In a 5G core network, a large number of entities perform processing tasks that need to be controlled and monitored. Whereas a 4G core might use tens of servers to run a set of network elements, a 5G network could employ thousands of containers. That step change means an exponential change in the number of interactions between these entities, making 5G observability much more complex.

Developed using open source technologies, HPE's observability platform employs a common set of tools across the four different layers of a 5G core network - the host level (the CPU, the memory and other infrastructural components), the virtualization layer, which collects container-related data, the network functions and the slice level. These tools can give the operator a precise view application-by-application, network function-by-network function, of all the key performance indicators and all the metrics related to each application, according to HPE.

HPE employs a network function management function (NFMF) to normalise the observability data flow from the various network functions, and expose this data via standard APIs. These APIs can then be subscribed and consumed by the OSS orchestration and assurance layers to perform actions, such as closed loop healing.

To identify the root cause of any problems, HPE employs artificial intelligence within the two new functions. The network data analytic function (NWDAF) can detect patterns of traffic, do trend analysis and then predict the condition in the network. As the fast control loop, the NWDAF can act directly on the network itself to redirect the traffic or increase the capacity. The management data analytic function (MDAF) uses machine learning to analyse information collected in a data lake and detect underlying problems in the network. When a problem emerges, these functions will probably have already detected it and have identified the root cause. The network is then reconfigured to resolve the issue.

As they are co-located with the network, the NFMF, NWDAF and MDAF can react faster than a traditional element manager system (EMS). By providing more granular and more sophisticated visibility, these new network functions enable a 5G operator to employ automated management tools, and ultimately develop a network that is able to manage itself.

Introduction

Cloud-native 5G networks will generate far more operational data than their predecessors. Massive volumes of valuable insights must be collected efficiently, integrated with operational support systems (OSS) and correlated across network functions.

This paper explores why cloud-native 5G networks need to be observable, what challenges need to be overcome and what kinds of tools are involved. It also considers how operators can translate the information collected by observability systems into action.

Developed by Mobile World Live in partnership with Hewlett-Packard Enterprise, the paper begins by defining observability and the potential benefits. It then goes on to consider how to apply observability to the highly-distributed and granular architecture of a 5G core network. Two further sections focus on how to integrate cloud-native observability tools into a telecoms environment and how to identify the root causes of any problems.

What is observability?

Now widely adopted, cloud computing is both abstract and virtualised. That is driving demand for a new breed of sophisticated tools that enable organisations to observe in real-time how their cloud-based IT systems are performing. Research firm IDG Connect defines observability as going beyond the core pillars of "IT monitoring, auditing, management and systems administration" to provide organisations with a clear real-time view of what is going on. "This is an area of deeper software log aggregation and analytics to enable us to really know how systems are working 'in flight' as they run in live operations," IDG notes.

As much of the data being generated in IT systems' log files will be raw and unstructured, observability platforms need to be able to analyse and correlate logs, metrics and traces to extract relevant signals, which can then be viewed through a visualisation layer to monitor system health against service level indicators and key performance and risk indicators.

Why observability is important for IT systems

At a high level, effective observability can give organisations greater confidence in running their IT in a cloud environment. It enables them to feel in control, rather than being semi-detached from the software and systems running their operations.

Once they have observability, organisations can immediately see the impact of changes in volumes or the introduction of new software on the system, while also detecting bugs, security breaches, and application and infrastructure performance bottlenecks. An observability platform should provide a continuous view of how systems are performing and any faults that are occurring.

At the same time, the observability concepts of monitoring, tracing, and visibility are a key enabler of DevOps – the concept of combining development and operations to efficiently and quickly deploy new software.

Like other complex IT systems, cloud native 5G core networks need to be observable to enable operators to check that they are performing as expected. Cellular networks are vulnerable to both physical disruptions and capacity issues. Furthermore, deep and accurate observability will help operators to automate 5G networks, and thereby improve both efficiency and performance.

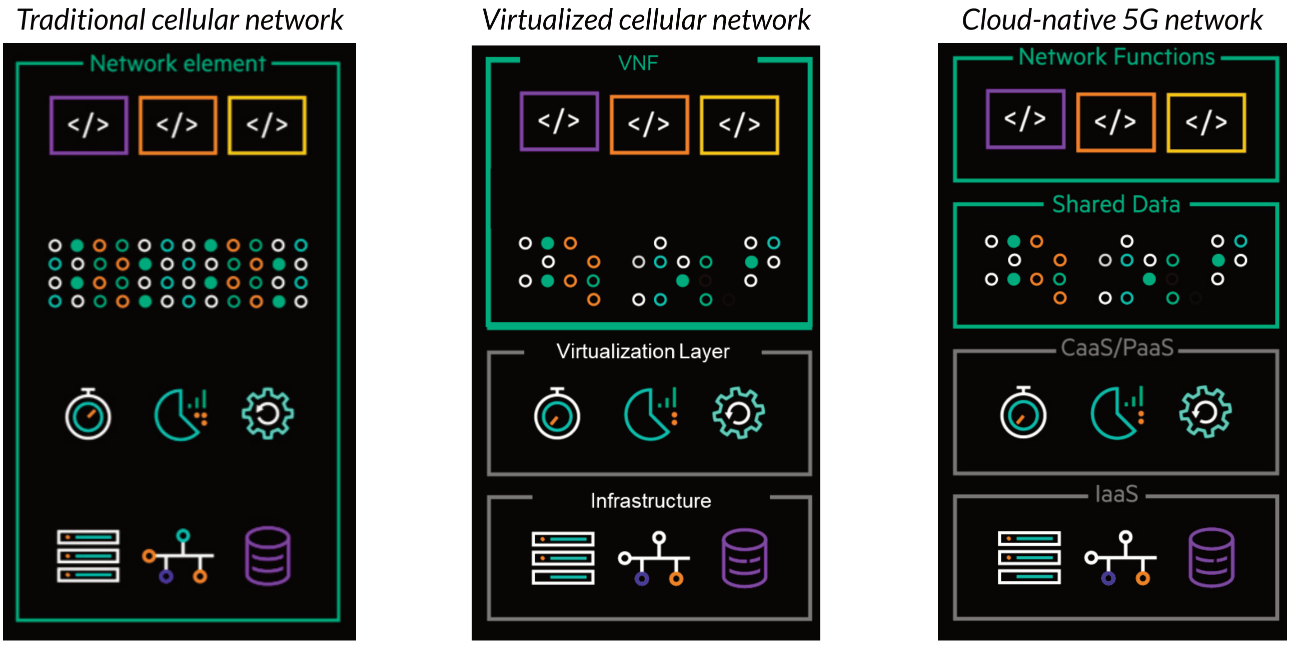
The challenges of observability in a 5G network

More entities, more interactions

Designed to support a much broader range of use cases, 5G core networks are very different from their predecessors. From an observability perspective, one of the key differences is the granularity of the entities performing network functions and processing tasks. In previous generations of cellular technologies, most of the network functions were delivered by monolithic boxes, which each performed a wide variety of different tasks (see the left hand pane on Figure 1).

During the first stages of virtualisation, this architecture evolved into a set of virtual machines, with each performing a distinctive task among the responsibilities of that network function (the centre pane on Figure 1). However, in a cloud native 5G core, the network functions are completely decomposed into a large number of smaller entities, called microservices (the right hand pane of Figure 1).

Figure 1: 5G introduces a much higher degree of granularity



Source: HPE

This new architecture is designed to be far more flexible and scalable, but it means a 5G operator will have a large number of entities performing processing tasks that need to be controlled and monitored. "In other words, 5G observability has the challenge of scale: there are a much larger number of entities performing network functions and interactions," explains Eduardo Ruano, 5G Core Stack, Outbound Product Manager, HPE Communications Technology Group. "While in 4G or in previous generations, a set of different network elements could run on tens of servers, we transitioned into having hundreds of virtual machines in what could be a 4G virtualised core. But now, with 5G, for a similar deployment, we're going to talk about thousands of containers."

The step change in the number of core network entities means an exponential change in the number of interactions between these entities, making 5G observability much more complex.

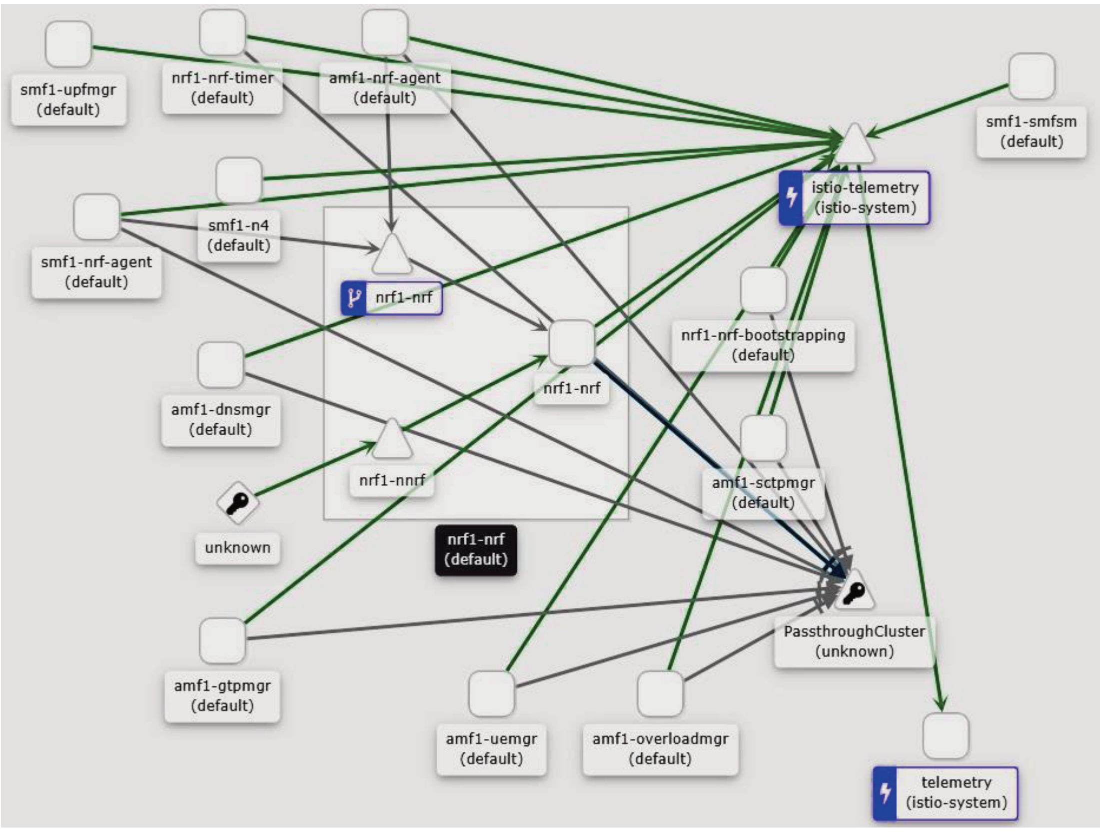
Figure 2 is a screenshot of the microservices interactions between three network functions in HPE's 5G core stack in a real deployment in its labs. It shows a particular slice in which there are three network functions: the AMF (access and mobility function), the SMF (session management function) and the NRF (network repository function). "Even with just three network functions in a very small scale and a very small capacity, it is still complex," notes Eduardo Ruano.

Extracting meaning from data

As well as monitoring each microservice and the interactions between them, an observability system needs to be able to translate the data it is collecting into meaningful information.

To complicate matters further, each network slice, made up of multiple network functions, is going to be deployed across a variety of data centres. As a result, there will be many points from which the operator needs to extract information and compile it and process it to extract meaning.

Figure 2: A 5G microservice mesh can be very complex



Source: HPE

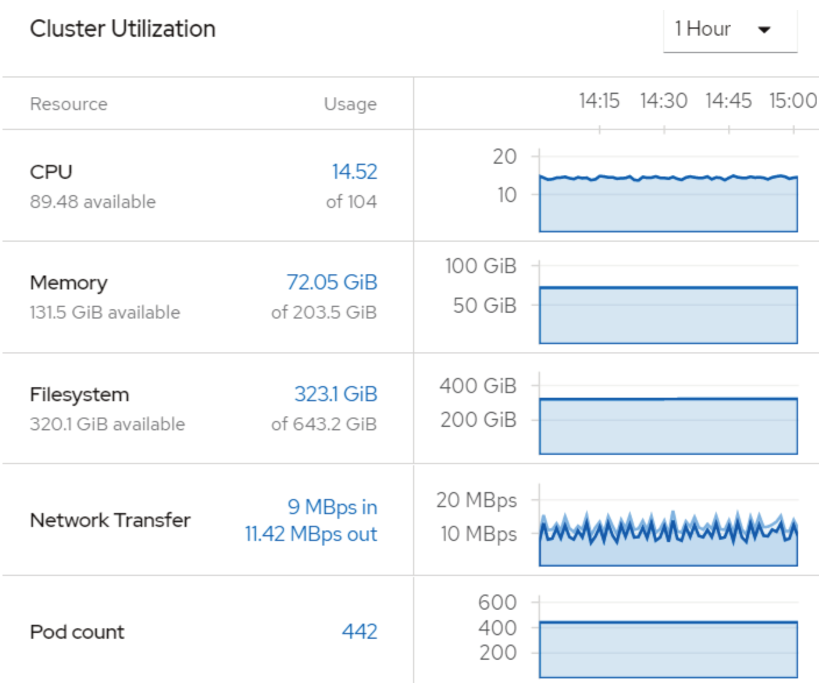
How to approach observability in a 5G core

An observability platform for a 5G core network can be divided into layers (as shown in the table below). The bottom layer is the host level, where there is the CPU, the memory and other infrastructural components. On top of that is the virtualization layer, which collects container-related data. The next layer contains the network functions - the application themselves. This layer generates information about the number of transactions, the number of authentication requests and other data. At the top is the slice information, which aggregates the system data at the slice level. HPE employs a common set of tools across the four different layers, developed using open source technologies.

Scope	KPIs/Metrics
Slices	QoS
Network functions	TS 28.552 Registration, Authentication, Mobility, PDU sessions
Cluster and containers	Pod, CPU, Memory, Network, Volumes
Host (hardware, IaaS, OS)	Server, CPU, Memory, Network, Storage, IOs

Figure 3 shows a screenshot of live metrics from the infrastructure layer, such as CPU usage, memory and networking.

Figure 3: A screenshot showing monitoring of the host layer

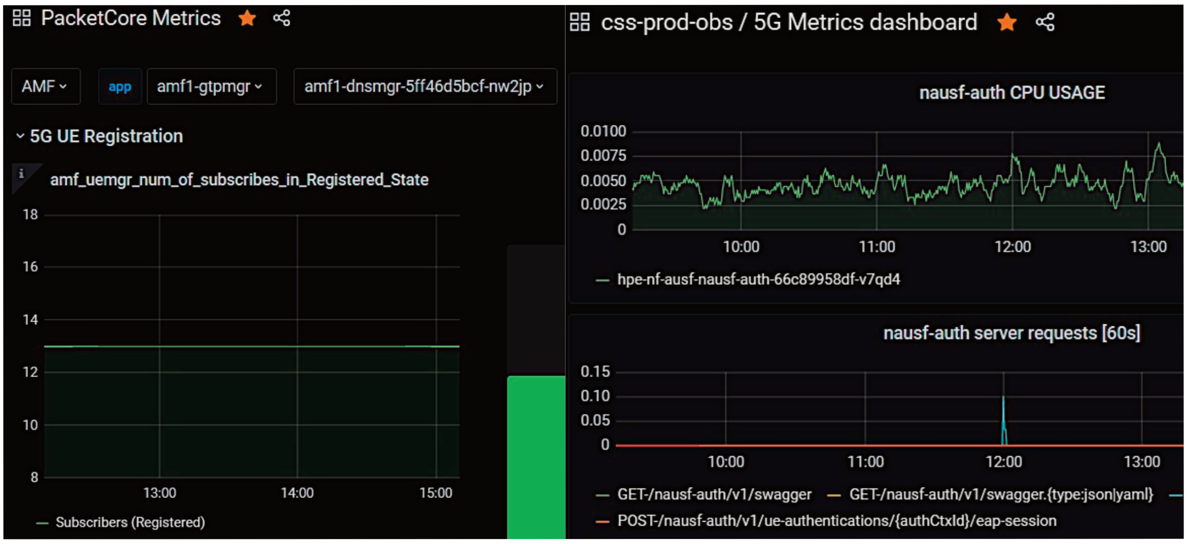


Source: HPE

HPE says its observability tools give the operator a precise view application-by-application, network function-by-network function, of all the KPIs and all the metrics related to each application. Figure 4 shows screenshots from a system monitoring the network functions. It shows performance metrics for the AMF (access and mobility function) – the number of subscribers registered on the network and authorisation activity.

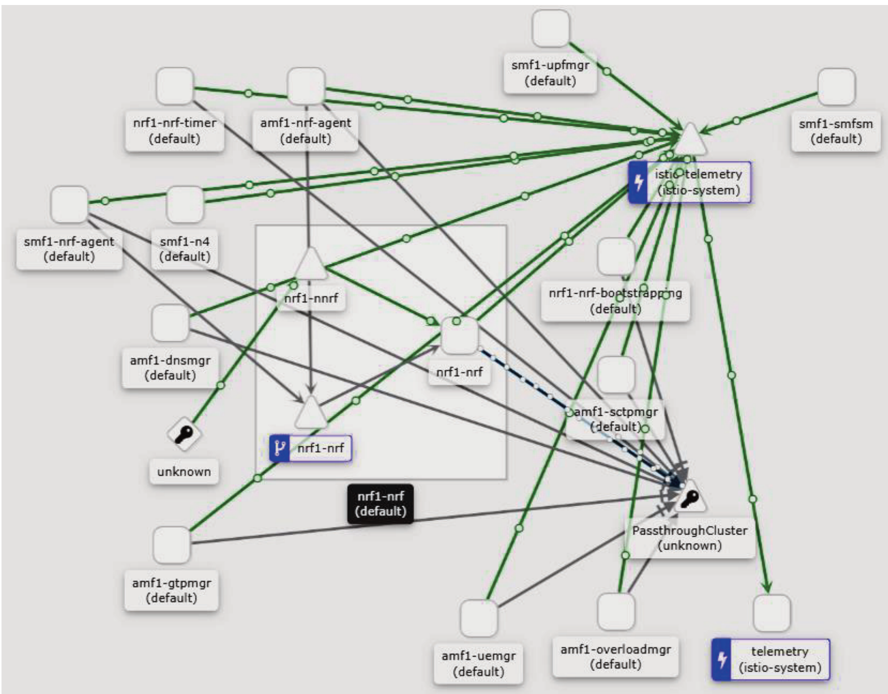
To monitor the mesh topology and all the interactions across the distributed microservices, HPE uses the open source tool Kiali. Figure 5, a screenshot from a monitoring system, shows the traffic (marked in green) between the nodes, along with potential errors (typically marked in red), which could be an unreachable element or HTTP for not-found resources, for example. In practice, this visualisation is animated to give the viewer a clear sense of what is happening. HPE says errors will be clearly visible as an alarm, enabling the operator to start drilling down to understand the root cause of the problem.

Figure 4: Screenshots showing live monitoring of network functions



Source: HPE

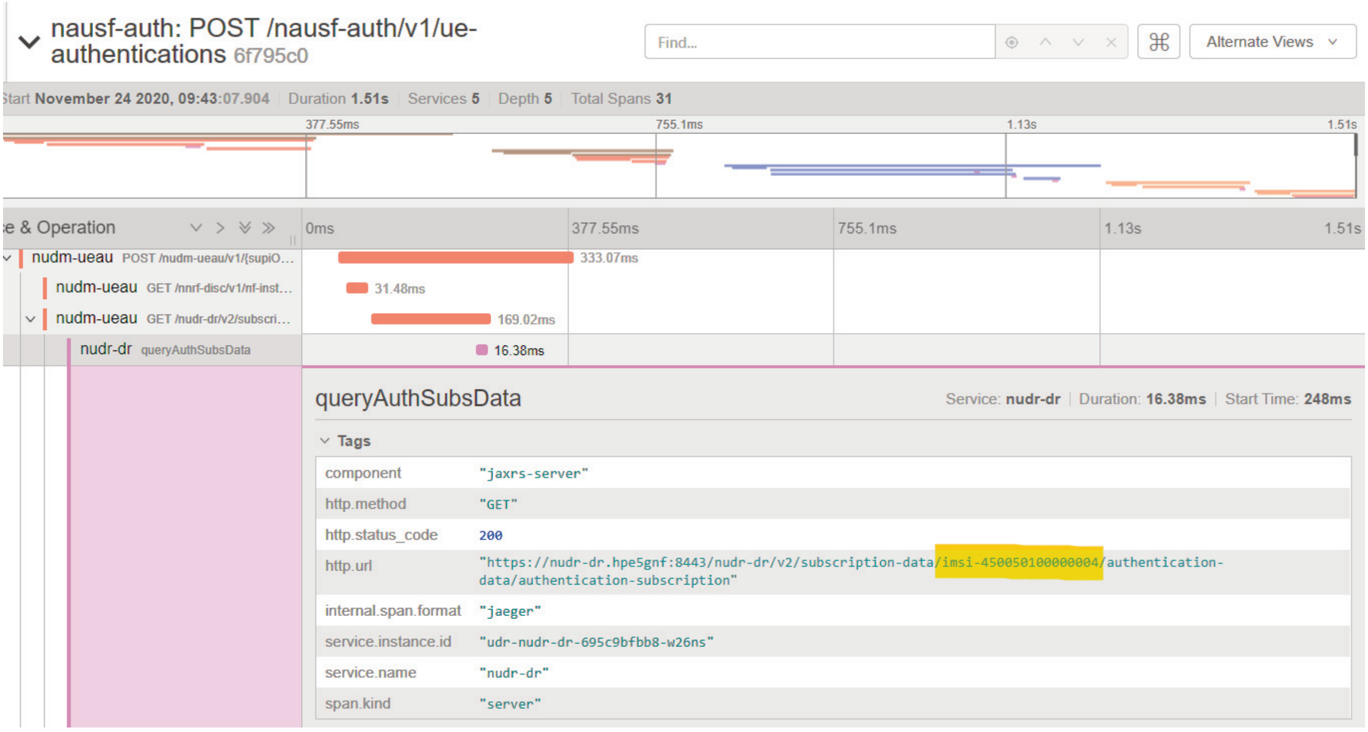
Figure 5: An observability tool can show the points of failure in a mesh



Source: HPE

The distributed architecture and intertwined services in a 5G network make point-to-point troubleshooting extremely difficult to use. For example, a basic operation, such as a user equipment registration, results in tens of interactions between microservices. Clearly, it is not practical to troubleshoot by collecting traces separately at each interface and then correlate them to investigate the problem and the root cause. Instead, HPE uses the open source tool Jaeger (see Figure 6) to collect and visualise entire spans of interactions between the microservices involved in a transaction.

Figure 6: A screenshot of the Jaeger tool monitoring authentication of user equipment



Source: HPE

Integrating cloud-native observability tools with telecoms systems

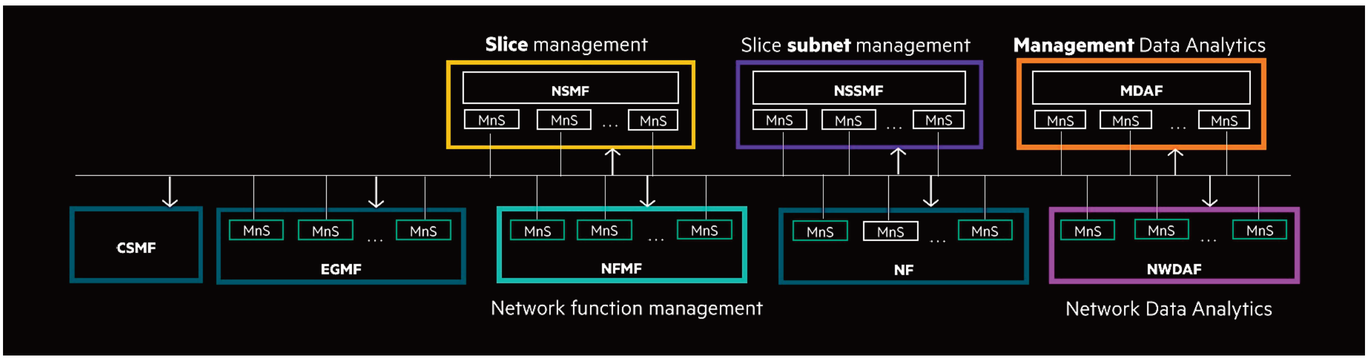
Element manager systems (EMS) have always played a pivotal role in data collection in cellular networks. With a cloud native 5G core, the EMS is superseded by multiple functions (see Figure 7), which bring the management capabilities closer to the network functions themselves.

While the configuration management of network functions can be integrated directly with an OSS or VNF (virtual network function) manager using a cloud native package automation, such as helm, the network function fault and performance data needs to be normalised, processed and exposed in a way that can be consumed by the telco OSS layer and even internally for other 5G network functions. "This is where the network function management function (NFMF) comes into the picture," explains Sebastien Klahr, 5G Core Stack Product Manager at HPE.

HPE's approach is to introduce a NFMF entity focusing on fault and performance, leaving the lifecycle management and the configuration management as a direct integration between the VNF, the OSS and the PaaS (platform-as-a-service). "The NFMF function bridges the cloud native 5G core network programme management with telco OSS in a 3GPP-compliant way and here it takes the observability data flow from the various network functions, normalises them and exposes the data in a specific telco fault and performance over standard APIs," says Sebastien Klahr. These APIs can then be subscribed and consumed by OSS orchestration and assurance layers to perform actions, such as closed loop healing.

Figure 8 provides an overview of HPE's approach to observability. The network is depicted at the bottom of the graphic, with the assurance system on the top right and the orchestration system on the top left. The network functions produce observability data, which is then aggregated by the NFMF and transmitted to the orchestration and assurance systems using APIs and alerts.

Figure 7: EMS has been decomposed into multiple functions



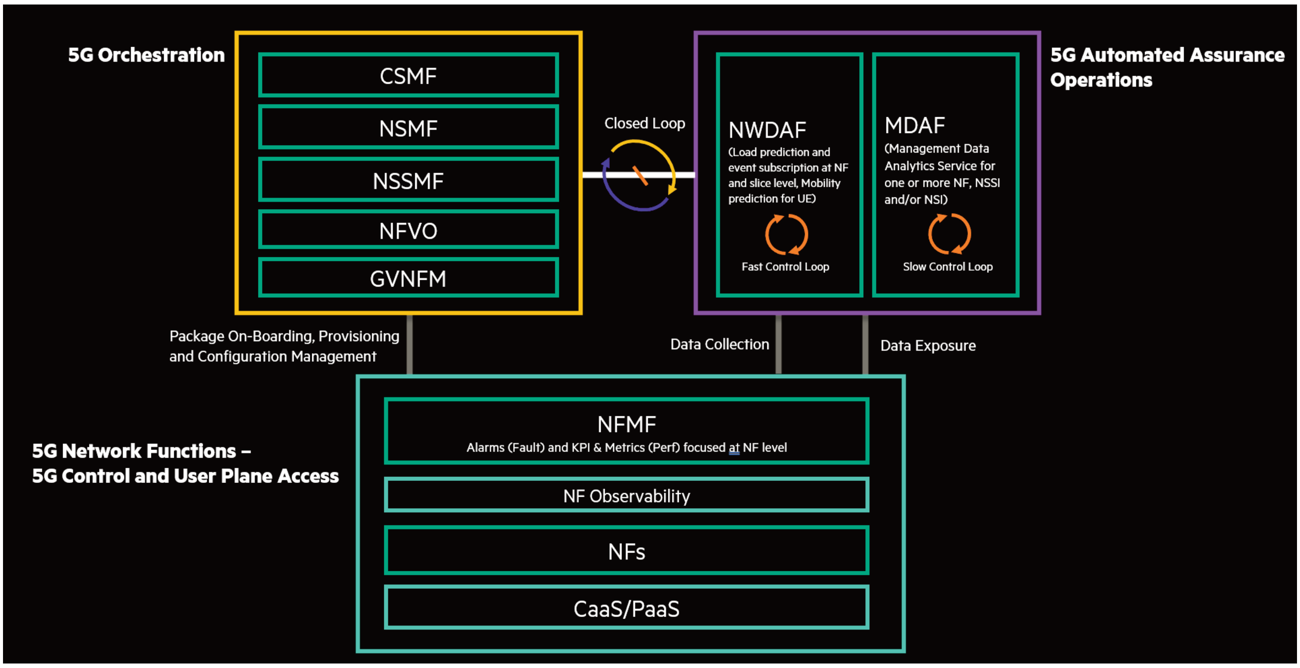
Source: HPE

The NFMF works with two new functions (introduced in the 3GPP standards and shown in Figures 8 and 9) to deliver full observability in a 5G core network:

1. The network data analytic function (NWDAF), which collects data and provides analytic services, in support of automation, using a request or subscription to network functions, enabling a fast control loop. It is similar to an embedded intelligent engine that is able to take actions on network behaviour. But the NWDAF will take immediate action in the event a threshold is crossed, for example.

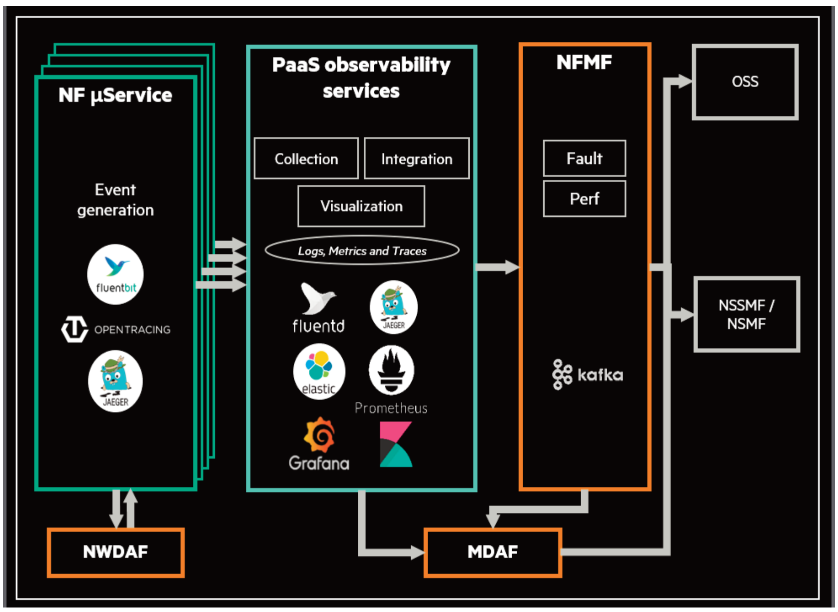
2. The management data analytic function (MDAF), which provides additional analytics features and enables the implementation of a slow control loop at a higher level in the management plane. The data is analysed to detect patterns and take appropriate actions, such as scaling out, due to a recurring increase of traffic.

Figure 8: How the NFMF works with the orchestration and assurance systems

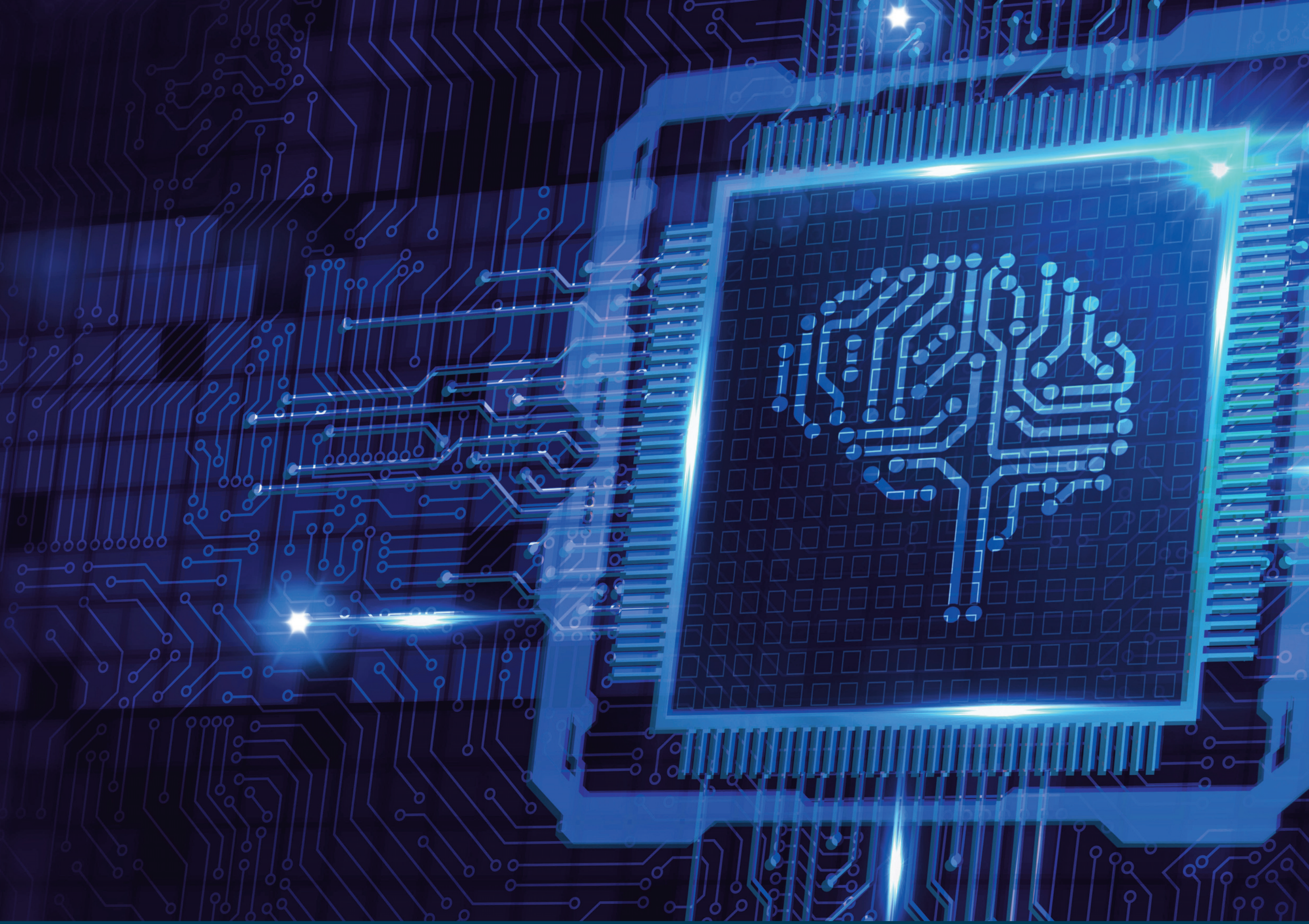


Source: HPE

Figure 9: HPE employs open source tools in the new management functions



Source: HPE



Identifying root causes

The NFMF cannot identify the root cause of any problems. To do that, HPE employs artificial intelligence/machine learning within the two new functions: the NWDAF and MDAF. "NWDAF is using pretty basic artificial intelligence algorithms to detect trending," explains Pierre Lavillat, 5G Core Stack & 5G Global Practice Manager at HPE. "If you have a pattern of traffic, where you know every day at 8pm, you have a surge of traffic for a specific service, this machine learning can do trend analysis and then predict the condition in the network." As the fast control loop, the NWDAF can act directly on the network itself to redirect the traffic or increase the capacity.

"Then when it comes to deep root cause and service impact analysis, here you have the MDAF, which is using machine learning, but on a more sophisticated level using pattern discovery in all the observability data," adds Pierre Lavillat. "We use this to train the algorithm to detect problems on the network. This is done using a data lake: we collect data, we do analysis and we detect patterns of problems."

As a result, when a problem emerges in a network, there is a high chance that the machine learning will have already detected it and have identified the root cause. It can then take remedial action to reconfigure the network and change the configuration to resolve the issue. In this way, HPE enables closed loop automation.

Conclusions

With potentially thousands of microservices in play, a 5G core network will generate vast amounts of data. The volumes of data will be far too much for human beings to process and act upon. As a result, the management of these networks needs to be done by automated systems underpinned by machine learning and artificial intelligence.

Standards body 3GPP anticipated this problem in advance and defined new management functions in the network: NFMF, NWDAF and MDAF. "They are not really replacing the EMS," notes Pierre Lavillat. "They are expanding the capability of what you could expect from an EMS because they are much more sophisticated: they will bring some of the capabilities that were done only at OSS level, but they will be co-located with the network and being closer to the network, they can react faster and they can directly interface and reconfigure the network." Although most 5G operators will continue to use an OSS, it will act more as a control tower that comes into play when something goes really wrong.

As a result of the increased granular visibility enabled by the new network functions, an operator can employ automated tools to manage a 5G network, and ultimately develop a network that is able to manage itself. "It will be able to do self-healing, self-scaling and be able to react to issues or traffic that are unexpected and will be much more efficient," concludes Pierre Lavillat.

HPE has over 30 years of experience in the telecoms industry, with more than 300 telco customers across 160 countries. In the core, more than 700 million subscribers across more than 80 carriers depend on HPE Mobile Core software. HPE's open telco solutions help operators evolve their networks and services to a 5G ready, cloudnative, servicebased architecture. As the edgetocloud platformasaservice company, our experience in hybrid cloud allows us to bring the cloud transformation and secure, carriergrade, standardsbased infrastructure to telecommunications networks.

Hewlett Packard Enterprise is the global edgetocloud platformasaservice company that helps organizations accelerate outcomes by unlocking value from all of their data, everywhere. Built on decades of reimagining the future and innovating to advance the way people live and work, HPE delivers unique, open and intelligent technology solutions, with a consistent experience across all clouds and edges, to help customers develop new business models, engage in new ways, and increase operational performance.

Learn more at hpe.com/info/5G

Mobile World Live is the premier destination for news, insight and intelligence for the global mobile industry. Armed with a dedicated team of experienced reporters from around the world, we are the industry's most trusted media outlet for breaking news, special features, investigative reporting, and expert analysis of today's biggest stories.

We are firmly committed to delivering accurate, quality journalism to our readers through news articles, video broadcasts, live and digital events, and more. Our engaged audience of mobile, tech and telecom professionals, including C-suite executives, business decision makers and influencers depend on the unrivalled content and analysis Mobile World Live provides to make informed business decisions every day.

Since 2016, Mobile World Live has also had a team of in-house media and marketing experts who work directly with our brand partners to produce bespoke content and deliver it to our audience in strategic yet innovative ways. Our portfolio of custom work - including whitepapers, webinars, live studio interviews, case studies, industry surveys and more – leverage the same level of industry knowledge and perspective that propels our newsroom.

Mobile World Live is published by, but editorially independent from, the GSMA, producing Show Daily publications for all GSMA events and Mobile World Live TV – the award-winning broadcast service of Mobile World Congress and home to GSMA event keynote presentations.

Find out more at www.mobileworldlive.com