RADCOM

Subscriber analytics and end-to-end troubleshooting for 5G networks

The information, content or material herein is provided "AS IS", is designated confidential and is subject to all restrictions in any law regarding such matters, and the relevant confidentiality and non-disclosure clauses or agreements issued prior to and/or after the disclosure. All the information in this document is to be safeguarded and all steps must be taken to prevent it from being disclosed to any person or entity other than the direct entity that received it directly from RADCOM. The text and drawings herein are for the purpose of illustration and reference only.

RADCOM reserves the right to periodically change information that is contained in this document; however, RADCOM makes no commitment to provide any such changes, updates, enhancements, or other additions to this document to you in a timely manner or at all.

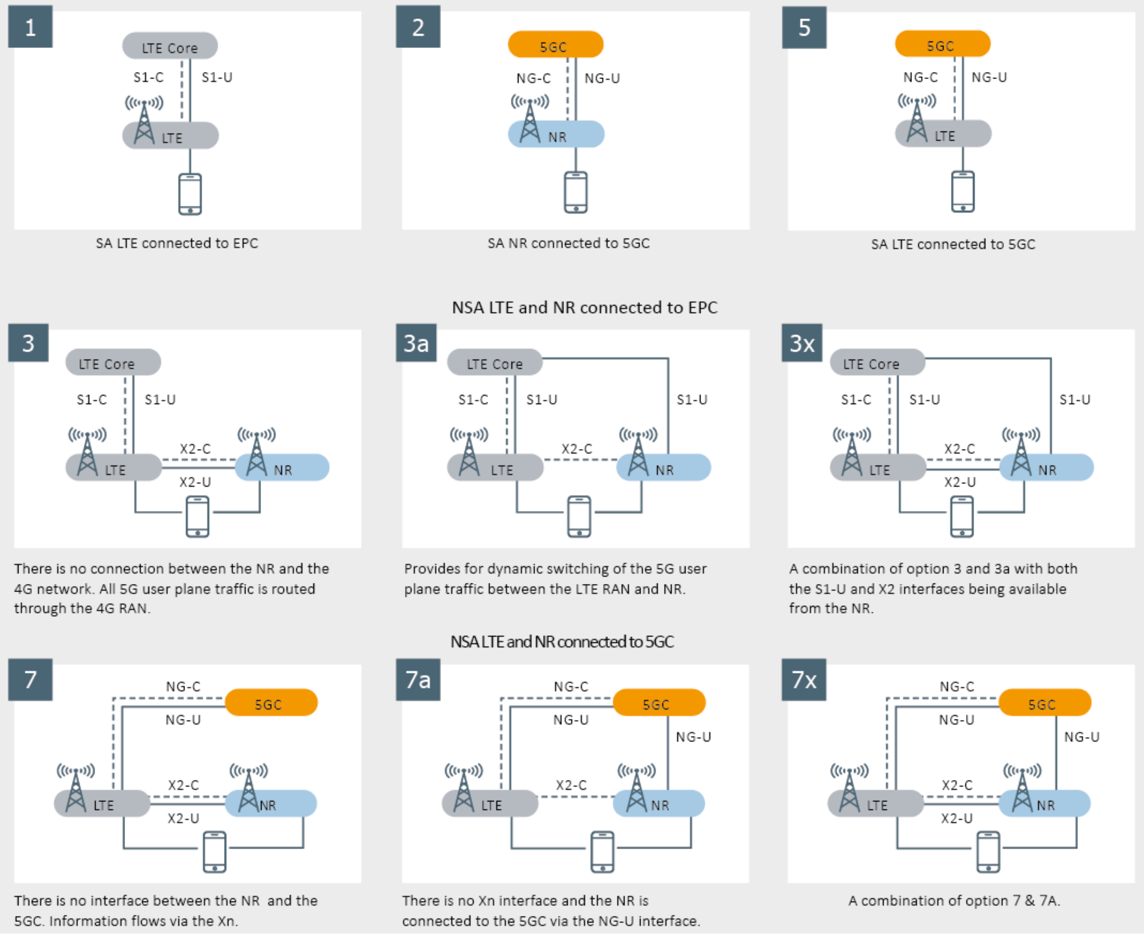
Publication Date: October 2020
Website:
http://www.radcom.com

Table of Contents

Introduction

With the introduction of 5G into a broad range of markets worldwide, the Fourth Industrial Revolution has begun. 5G promises a future in which people are always connected. Technology is deeply embedded in society, providing high speed and instant connectivity for any device and any person anywhere on earth, without restrictions.

Previous generations of mobile networks were purpose-built for delivering communication services such as voice (2G), mobile data services like messaging (3G), and then mobile broadband (4G) for multimedia on the move with services such as video chat and video streaming. 5G is much more than the next iteration of mobile networks.
In the initial stages, 5G will deliver enhanced mobile broadband with higher data speeds (up to twenty times that of 4G) and better coverage for use cases such as Fixed Wireless Access (FWA). However, over time, 5G will transform the role that telecommunications play in society. 5G will enable new use cases such as autonomous cars, remote control of critical infrastructure/machinery, smart-grid control, industrial automation, robotics, drone control, and remote telehealth services, which will be empowered by an ultra-reliable, low latency network that will revolutionize our lives.



SA LTE connected to EPC

SA NR connected to 5GC

SA LTE connected to 5GC

NSA LTE and NR connected to EPC

There is no connection between the NR and the 4G network. All 5G user plane traffic is routed through the 4G RAN.

Provides for dynamic switching of the 5G user plane traffic between the LTE RAN and NR.

A combination of option 3 and 3a with both the S1-U and X2 interfaces being available from the NR.

NSA LTE and NR connected to 5GC

There is no interface between the NR and the 5GC. Information flows via the Xn.

There is no Xn interface and the NR is connected to the 5GC via the NG-U interface.

A combination of option 7 & 7A.

Figure 1 - 5G deployment options

Initial, 5G networks are being deployed in non-standalone mode (NSA) with operators using an upgraded 4G core network and the new 5G radio (5G NR) that is often virtualized. Many operators have begun deploying their cloud-native 5G core (5GC) to deliver more advanced 5G services in standalone (SA) mode. This cloud-native core will enable operators to reduce costs, accelerate the introduction of new services, and deliver faster iterations to their network. Creating a dynamic, open, scalable, and modular platform to deliver the next iteration of exciting mobile innovations.

Challenges for operators

A significant promise of 5G is the improvement in the customer experience. As a result, the end-to-end service quality and customer experience will be of supreme importance to operators. However, this will require the right service assurance approach to be implemented by operators to gain the analytics to produce relevant insights from these complex networks, with billions of possible dimensions to the customer experience (new types of devices, locations, network topologies, OTT services, and consumers). With 5G, the experience is everything and the foundation for new business use cases and critical revenue streams.

Delivering the expected customer experience

Traditionally, telecom operators have faced challenges in delivering the expected customer experience. An Analysys Mason customer survey[1] showed that telecom operators in Europe and North America had Net Promoter Scores (NPS) between –5 and 40 compared to companies like Amazon and Netflix, scoring 50 or above. Additionally, in the Temkin Experience report, it was found that telecom operators in the US consistently ranked at the bottom end of customer satisfaction.[2]

CSPs must recalibrate their customer experience strategies to tackle the increasing network complexity as they embark on various network transformation initiatives such as 5G.
Anil Rao, Principal Analyst, Analysys Mason

The foundation for 5G success is focusing on the customer experience. Even more than past mobile technology iterations, operators that will fail to understand the customer experience and will not smartly monitor and rectify customer-affecting service issues will suffer churn. Continually ensuring a high quality of experience (QoE) and service (QoS) has never been easy, and 5G will make this even more challenging. With a cloud-native architecture being essential for 5G, operators will need to master network virtualization with its new architecture and dynamic nature to lay down the foundations for their next-generation services.
Despite this, all these underlaying changes will need to be transparent to the operators' customers, and the operator needs to ensure the QoE and QoS, however intricate the situation. Furthermore, as the 5G rollout continues, operators will need to support the new use cases around ultra-low latency, edge clouds, and network slicing. In contrast, the number of devices and the amount of traffic that needs monitoring continues to rise dramatically. Suppose operators are to transition to becoming DSPs successfully. In that case, they understand the need to change customer perceptions and reach the same level of customer experience and brand loyalty that web-first companies like Amazon, Netflix, and Google offer their customers.

[1] https://www.analysysmason.com/Research/Content/Reports/Mobile-satisfaction-Europe-USA-RDMM0

[2] https://www.qualtrics.com/docs/xmi/XMI_TemkinExperienceRatings-2018.pdf

Managing the complexity of cloud-native networks

Implementing a cloud-native 5G network brings many benefits to the operator in operational flexibility and scalability while laying down the foundations for the next generation of exciting, dynamic customer and business services. However, there are many challenges involved. The underlying architecture is more complicated than previous iterations of networks with hundreds of virtualized network functions deployed from the RAN to the network core. Up until recently, these functions were proprietary hardware solutions supported by Network Equipment Providers (NEPs) and known and deployed by network engineers for years. In the 5G core functions are virtual, dynamic, and can be launched on-demand with certain functions running alongside other functions on the same hardware, and therefore east-west and north-south traffic need to be monitored.

Operators integrate new management layers to streamline and automate their service deployments and manage services using a unified service level policy that will lead to a closed-loop network. Although the industry is moving to containerization and operators are building greenfield 5G networks, there remains a legacy network to manage for most operators. So, operators will need to understand the end-to-end service quality running across both network domains. Add to this the significant increase in data traffic. Millions of more devices connecting to the cloud operators have significant challenges in understanding the end-to-end customer experience and troubleshooting the network performance.

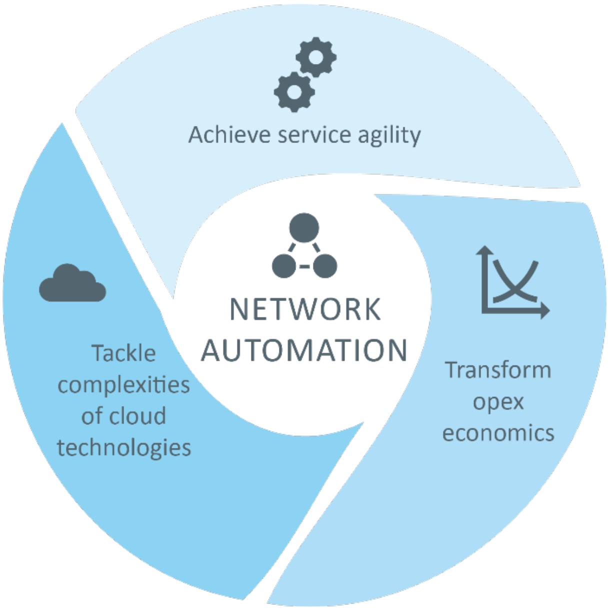


Figure 2 - Goals of operators' network automation

With 5G bringing faster data rates and early use cases like enhanced mobile broadband and fixed wireless access, it becomes even more critical for operators to overcome customer pain points to give them a seamless customer experience and successfully launch and optimize new quality services. As operators deploy their 5G, they must implement the correct assurance strategy.

The cornerstone to achieving real-time subscriber analysis

For 5G, operators must gain a service-level awareness and understand the customer experience through their subscribers' eyes. For example, is the subscriber trying to watch streaming video on a device that recently had its firmware updated and is now suffering from repeated buffering? In this case, monitoring the resource and network layer is not enough. To understand why the video is not streaming adequately, we need to see the end-to-end service layer.

Operators can do this by integrating a probe layer into their assurance strategy that allows them to understand the end-to-end service quality, includes real-time subscriber analytics, and provides troubleshooting tools. End-to-end, containerized probing enables the operator to;

- Enrich network analytics with real-time subscriber analytics to understand the end-to-end service quality and troubleshoot network degradations
- Integrate service assurance with network orchestration to drive closed-loop automation
- Deploy automated assurance that is containerized and controlled by Kubernetes to deploy and scale as part of the service lifecycle

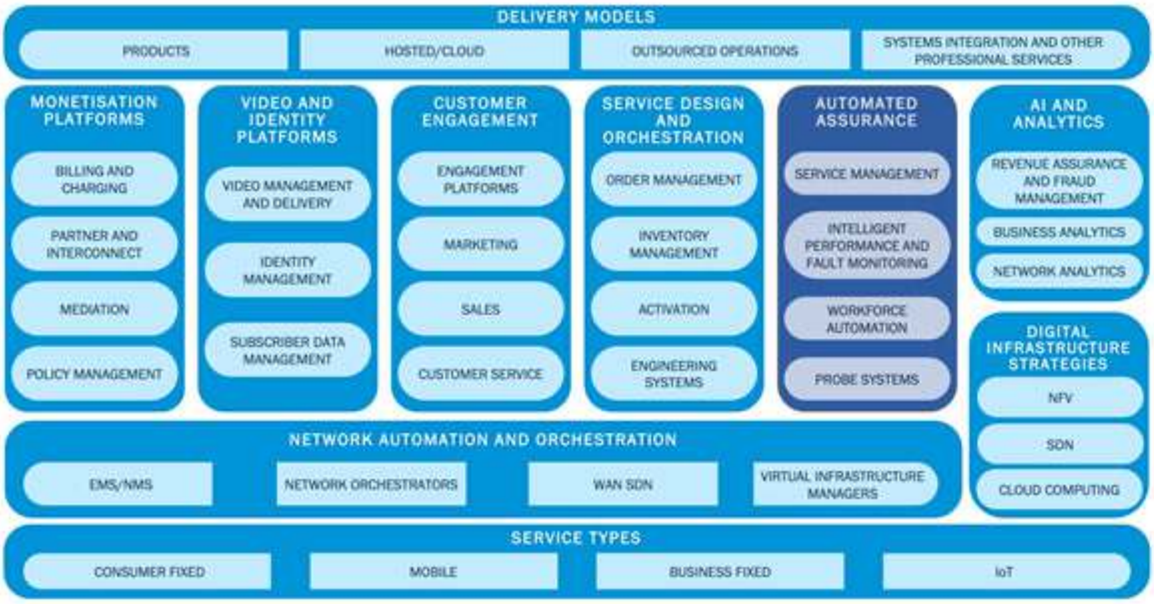


Figure 3 – Source: Analysys Mason - Telecoms automated assurance segmentation

With this strategy in place, operators will deliver a superior customer experience that will build brand loyalty, provide an edge over competitors, and enable operators to manage the complexity of a cloud-native 5G core.

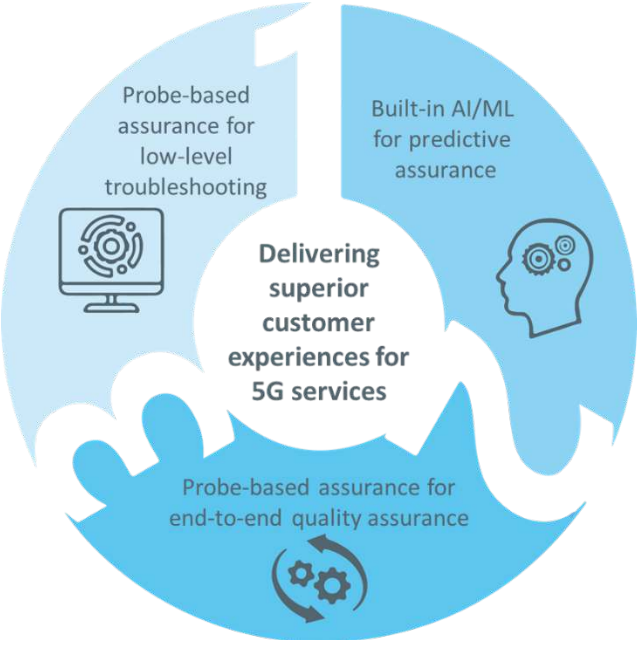
Probe-based assurance advantages

Probe-based assurance is key to gaining real-time subscriber and service analysis that provides an understanding of the customer experience, gives insights into the end-to-end service quality and is essential for troubleshooting new services.

"A probe-based approach is an essential part of the end-to-end a customer experience and service quality management solution."
TMForum

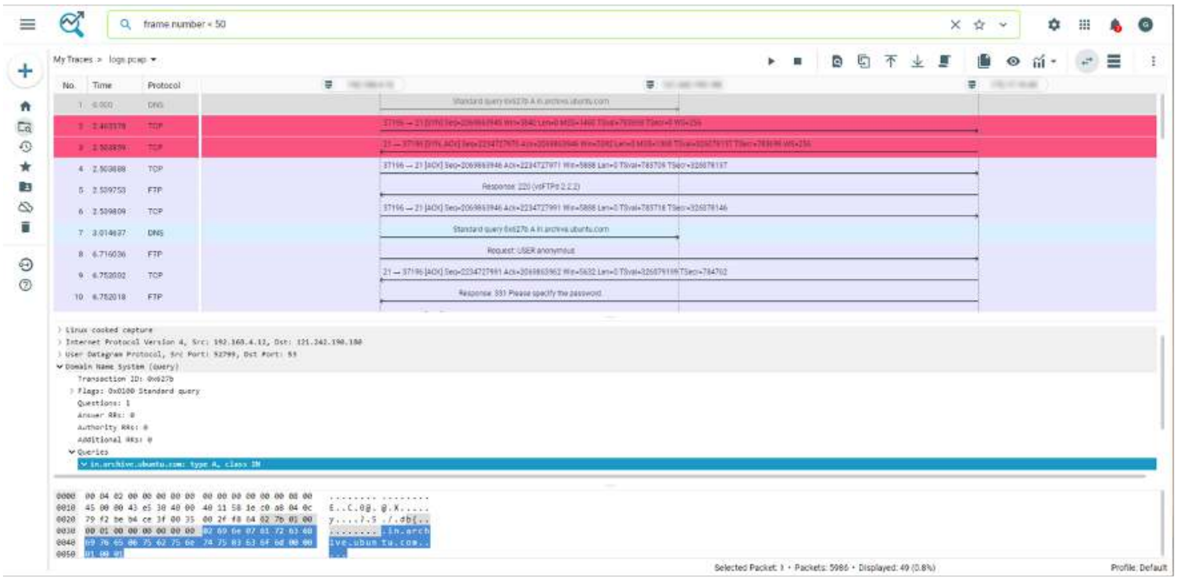
"Probes will play a vital role in supporting efforts to improve the customer experience."
Analysis Mason

When measuring the detail of a customer's end-to-end service performance, probes are the most effective way to gain real-time subscriber analytics and service performance insights. Probes enable real-time and historical end-to-end call tracing from the user device to the core network and generating XDRs (eXtended Detail Records) per session per subscriber that contains vital information about single-user sessions.


Probe-based assurance for low-level troubleshooting
Built-in AI/ML for predictive assurance
Delivering superior customer experiences for 5G services
Probe-based assurance for end-to-end quality assurance

XDRs are transmitted from the probing layer to a central backend processing and mediation layer where the XDRs are further processed (enrichment, correlation, aggregation, alarming) and inserted into a database. In parallel, raw packets are stored locally in the probe's storage and are available for packet-level tracing.

In essence, probes watch all the traffic that flows through the network and filter out individual transactions to compute the service quality experienced by each call or data transfer. They provide granular data that allows operators to determine the service quality at a per-service (QoS) and per-user (QoE) granularity across multiple transport technologies.



frame number = 50

My Traces > logs.pcap

```
No.  Time      Protocol
1    0.000     DNS    Standard query 0x627b A in.archive.ubuntu.com
2    2.463378  TCP    37196 → 21 [SYN] Seq=2069863945 Win=5840 Len=0 MSS=1460 TSval=783699 TSecr=0 WS=256
3    2.503839  TCP    21 → 37196 [SYN, ACK] Seq=2234727970 Ack=2069863946 Win=5792 Len=0 MSS=1368 TSval=326078137 TSecr=783699 WS=256
4    2.503688  TCP    37196 → 21 [ACK] Seq=2069863946 Ack=2234727971 Win=5888 Len=0 TSval=783709 TSecr=326078137
5    2.539753  FTP    Response: 220 (vsFTPd 2.2.2)
6    2.539809  TCP    37196 → 21 [ACK] Seq=2069863946 Ack=2234727991 Win=5888 Len=0 TSval=783718 TSecr=326078146
7    3.014637  DNS    Standard query 0x627b A in.archive.ubuntu.com
8    6.716036  FTP    Request: USER anonymous
9    6.752002  TCP    21 → 37196 [ACK] Seq=2234727991 Ack=2069863962 Win=5632 Len=0 TSval=326079199 TSecr=784702
10   6.752018  FTP    Response: 331 Please specify the password
```

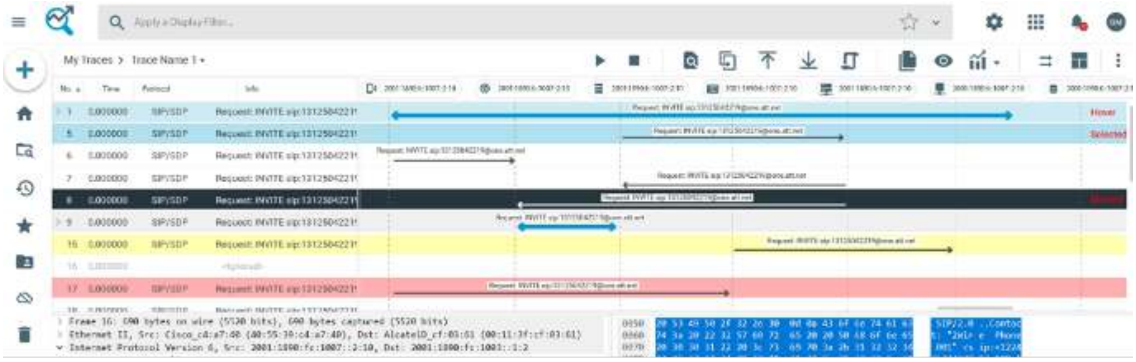
```
> Linux cooked capture
> Internet Protocol Version 4, Src: 192.168.4.12, Dst: 121.242.190.180
> User Datagram Protocol, Src Port: 52799, Dst Port: 53
v Domain Name System (query)
    Transaction ID: 0x627b
  > Flags: 0x0100 Standard query
    Questions: 1
    Answer RRs: 0
    Authority RRs: 0
    Additional RRs: 0
  v Queries
    v in.archive.ubuntu.com: type A, class IN
```

```
0000  00 04 02 00 00 00 00 00  00 00 00 00 00 00 08 00   ........ ........
0010  45 00 00 43 e5 30 40 00  40 11 58 1e c0 a8 04 0c   E..C.0@. @.X.....
0020  79 f2 be b4 ce 3f 00 35  00 2f f8 64 62 7b 01 00   y....?.5 ./.db{..
0030  00 01 00 00 00 00 00 00  02 69 6e 07 61 72 63 68   ........ .in.arch
0040  69 76 65 06 75 62 75 6e  74 75 03 63 6f 6d 00 00   ive.ubun tu.com..
0050  01 00 01                                          ...
```

Selected Packet: 1 • Packets: 5986 • Displayed: 49 (0.8%)   Profile: Default

Figure 4 - Probe-based data offers engineers the ability to view full decodes of captured packets

This real-time subscriber and service analysis provides a different dimension to the network-focused data. Network counters don't look at the customer experience and don't give the end-to-end picture, and there is no correlation. Therefore, the operator can't use this data to understand the end-to-end service quality. They don't have the raw trace that includes checking real-time and historical data per user at the packet level troubleshoot. Probes capture all signaling and user data events in the networks and integrate it into the operators' assurance solution.



Apply a Display Filter...

My Traces > Trace Name 1

```
No.  Time      Protocol  Info
1    0.000000  SIP/SDP   Request: INVITE sip:13125042211
5    0.000000  SIP/SDP   Request: INVITE sip:13125042211
6    0.000000  SIP/SDP   Request: INVITE sip:13125042211
7    0.000000  SIP/SDP   Request: INVITE sip:13125042211
8    0.000000  SIP/SDP   Request: INVITE sip:13125042211
9    0.000000  SIP/SDP   Request: INVITE sip:13125042211
15   0.000000  SIP/SDP   Request: INVITE sip:13125042211
16   0.000000
17   0.000000  SIP/SDP   Request: INVITE sip:13125042211
```

```
> Frame 16: 690 bytes on wire (5520 bits), 690 bytes captured (5520 bits)
> Ethernet II, Src: Cisco_c4:a7:40 (40:55:39:c4:a7:40), Dst: AlcatelD_cf:03:61 (00:11:3f:cf:03:61)
v Internet Protocol Version 6, Src: 2001:1890:fc:1007::2:10, Dst: 2001:1890:fc:1003::1:2
```

Figure 5 - Probe-based assurance enables deep-packet analysis and troubleshooting

Although network equipment providers (NEPs) can sometimes provide raw packets for specific users from network events/counters, it is presented in PCAP format and not a GUI, which means the engineer will need to use external tools such as Wireshark to troubleshoot, and this is limited to single-user access and not integrated into a cloud-based solution and the broader assurance solution.

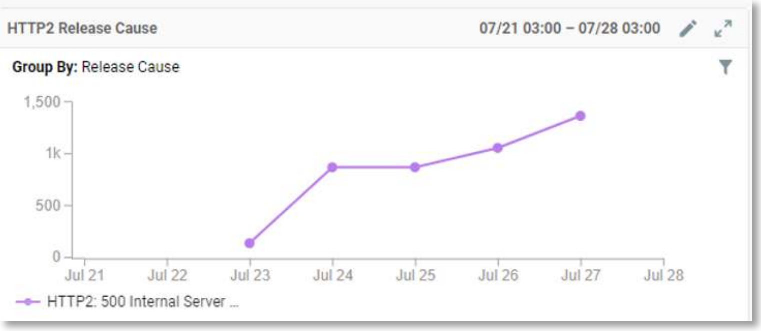
By utilizing probes as part of the assurance solution, operators can position real-time subscriber analysis at the center of their assurance strategy from top management to technical teams. Giving operators the capacity to:

- Analyze real-world network traffic and obtain reliable insight into the subscriber experience
- Analyze the performance and behavior of network components in a dynamic environment via KPIs and KQIs
- Generate XDRs (eXtended Detail Records) per session per subscriber to troubleshoot any issue on the network and identify its root cause

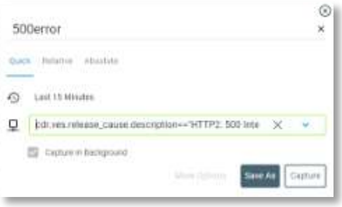
Troubleshooting a customer-affecting incident can be performed via a drill-down from the KPI layer right through to an individual network packet. Furthermore, user plane capture can be performed on-demand with access to all deployed probes and specific filters to capture the relevant data. This is then integrated into a BI or delivered as real-time streaming analytics fed into the operators' BSS/OSS or network orchestration.

An example of a troubleshooting workflow for 5G

The HTTP 500 Internal Server Error server error response code indicates that the server encountered an unexpected condition that prevented it from fulfilling the request. This error response is a generic "catch-all" response. Using probe-based assurance, operators can troubleshoot this error by drilling down to the packet level and viewing the call flow.



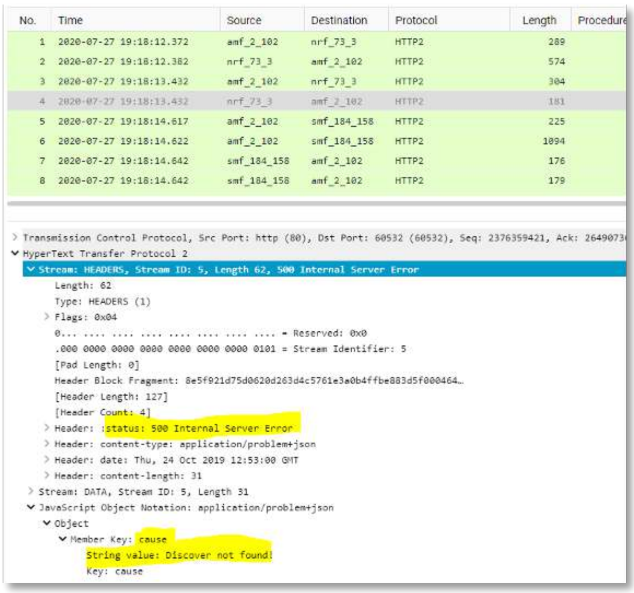
- A user filters for the HTTP 500 internal errors
- Opens the packet-based call flow in the packet analyzer


500error

Last 15 Minutes

cdr.ves.release_cause.description=="HTTP2: 500 Inte

- “If the discovery request fails at the NRF due to NRF internal errors, the NRF shall return "500 Internal Server



No.	Time	Source	Destination	Protocol	Length	Procedure
1	2020-07-27 19:18:12.372	amf_2_102	nrf_73_3	HTTP2	289	
2	2020-07-27 19:18:12.382	nrf_73_3	amf_2_102	HTTP2	574	
3	2020-07-27 19:18:13.432	amf_2_102	nrf_73_3	HTTP2	304	
4	2020-07-27 19:18:13.432	nrf_73_3	amf_2_102	HTTP2	181	
5	2020-07-27 19:18:14.617	amf_2_102	smf_184_158	HTTP2	225	
6	2020-07-27 19:18:14.622	amf_2_102	smf_184_158	HTTP2	1094	
7	2020-07-27 19:18:14.642	smf_184_158	amf_2_102	HTTP2	176	
8	2020-07-27 19:18:14.642	smf_184_158	amf_2_102	HTTP2	179	

```
> Transmission Control Protocol, Src Port: http (80), Dst Port: 60532 (60532), Seq: 2376359421, Ack: 2649073
v HyperText Transfer Protocol 2
  v Stream: HEADERS, Stream ID: 5, Length 62, 500 Internal Server Error
      Length: 62
      Type: HEADERS (1)
    > Flags: 0x04
      0... .... .... .... .... .... .... .... = Reserved: 0x0
      .000 0000 0000 0000 0000 0000 0000 0101 = Stream Identifier: 5
      [Pad Length: 0]
      Header Block Fragment: 8e5f921d75d0620d263d4c5761e3a0b4ffbe883d5f000464…
      [Header Length: 127]
      [Header Count: 4]
    > Header: :status: 500 Internal Server Error
    > Header: content-type: application/problem+json
    > Header: date: Thu, 24 Oct 2019 12:53:00 GMT
    > Header: content-length: 31
  > Stream: DATA, Stream ID: 5, Length 31
  v JavaScript Object Notation: application/problem+json
    v Object
      v Member Key: cause
          String value: Discover not found!
          Key: cause
```

- The error status code with the ProblemDetails IE provides details of the error.” (TS 29.510[3])
- ProblemDetails IE = “Discover not found!”

[3] https://www.etsi.org/deliver/etsi_ts/129500_129599/129510/15.01.00_60/ts_129510v150100p.pdf

Embedded DPI

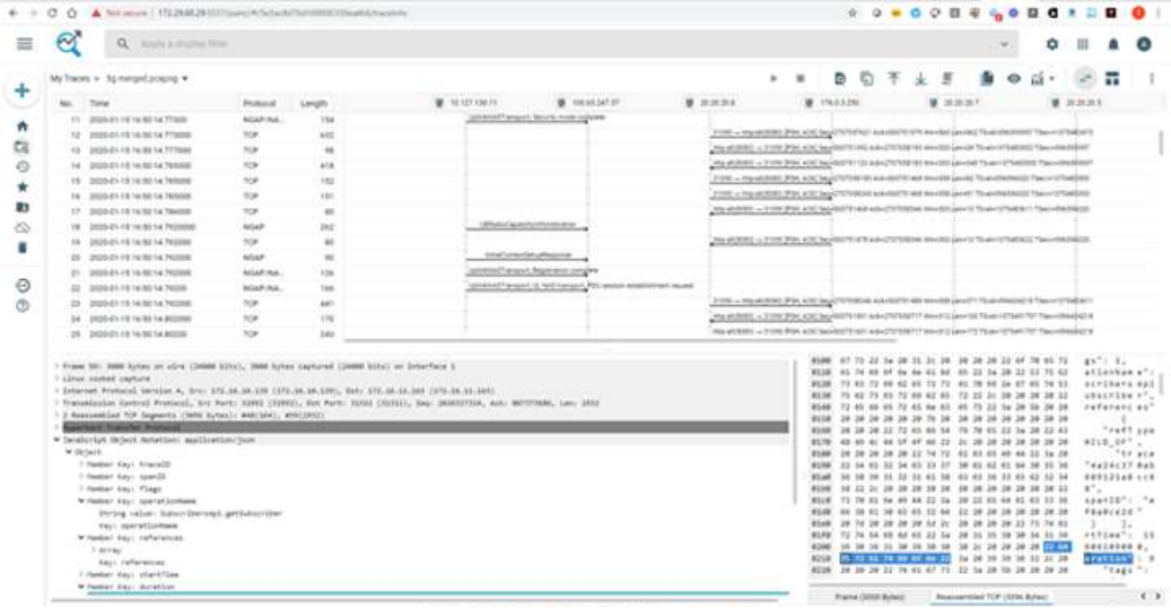


Figure 6 - Using RADCOM Packet Analyzer to display 5GC OpenTracing data

The introduction of 5G and increased complexity of networks will increase the need for probing based assurance. Access to actionable data is essential to enable automation and customer experience management and apply technologies such as Artificial Intelligence (AI) and Machine Learning (ML) to analyze the subscriber data. For example, DPI functionality can be embedded into the probes with machine learning algorithms powered by mass video streaming samples to identify non-encrypted and encrypted video traffic and provide insights into the quality of experience.

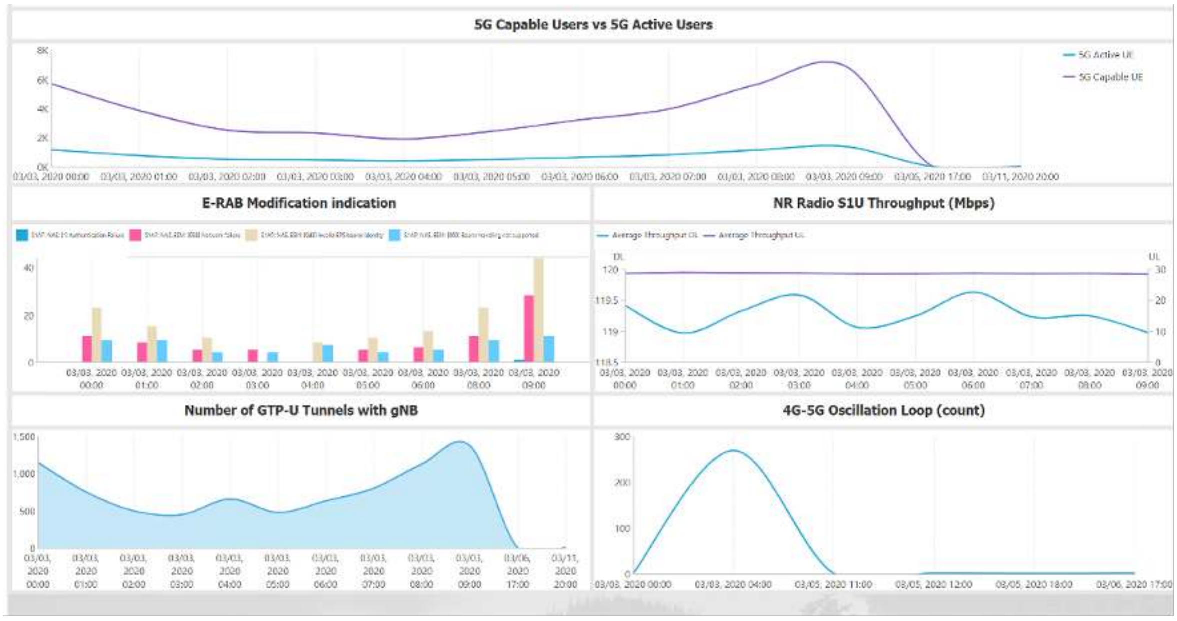


Figure 7 – Smartly monitoring 5G subscribers is critical for rolling out new 5G services

Probe-based data empowers the operator with comprehensive root cause analysis and troubleshooting tools at both a high and granular level, with drill-down capabilities to the packet or subscriber level. With so many new technologies, functions, and network architectural changes, both high-level and low-level tools in a container-based solution will be critical for operators transitioning to the cloud for NFV and 5G. However, for 5G, operators need to deploy a probe-based solution comprised of cloud-native functions (CNFs).

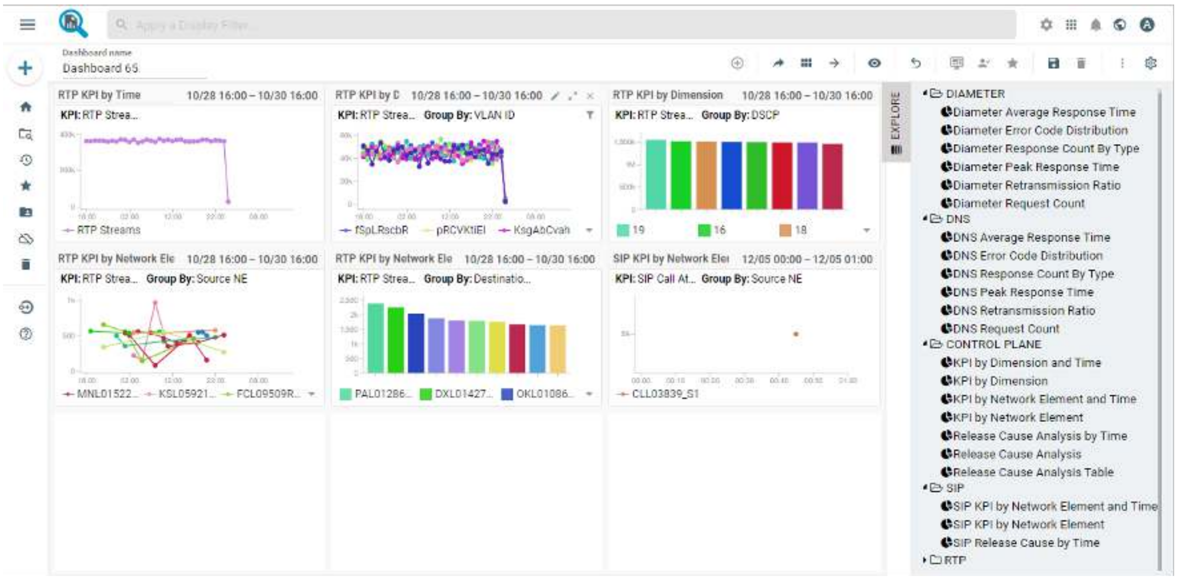

Apply a Display Filter...
Dashboard name
Dashboard 65
RTP KPI by Time 10/28 16:00 – 10/30 16:00
KPI: RTP Strea...
RTP Streams
RTP KPI by D 10/28 16:00 – 10/30 16:00
KPI: RTP Strea... Group By: VLAN ID
fSpLRscbR
pRCVKtIEI
KsgAbCvah
RTP KPI by Dimension 10/28 16:00 – 10/30 16:00
KPI: RTP Strea... Group By: DSCP
19
16
18
RTP KPI by Network Ele 10/28 16:00 – 10/30 16:00
KPI: RTP Strea... Group By: Source NE
MNL01522...
KSL05921...
FCL09509R...
RTP KPI by Network Ele 10/28 16:00 – 10/30 16:00
KPI: RTP Strea... Group By: Destinatio...
PAL01286...
DXL01427...
OKL01086...
SIP KPI by Network Elei 12/05 00:00 – 12/05 01:00
KPI: SIP Call At... Group By: Source NE
CLL03839_S1
EXPLORE
DIAMETER
Diameter Average Response Time
Diameter Error Code Distribution
Diameter Response Count By Type
Diameter Peak Response Time
Diameter Retransmission Ratio
Diameter Request Count
DNS
DNS Average Response Time
DNS Error Code Distribution
DNS Response Count By Type
DNS Peak Response Time
DNS Retransmission Ratio
DNS Request Count
CONTROL PLANE
KPI by Dimension and Time
KPI by Dimension
KPI by Network Element and Time
KPI by Network Element
Release Cause Analysis by Time
Release Cause Analysis
Release Cause Analysis Table
SIP
SIP KPI by Network Element and Time
SIP KPI by Network Element
SIP Release Cause by Time
RTP

Figure 8 – Monitoring KPIs across all domains, including mobile data, mobile voice, IMS, VoLTE, VoWiFi, OTT applications, fixed-line data, and fixed-line voice

Fully-containerized probes

CNFs are built using a containerized, microservices-based architecture, which provides the operator with many benefits when running their cloud network. Their microservices architecture means that each CNF comprises small independent processes or elements that all communicate and enable a modular approach to system building. A microservices architecture means the CNF is built-in parts, almost a Lego-like structure, removing and adding pieces as needed, keeping it lightweight and agile.

This means that containerized probes are automated and controlled with high granularity levels, meaning it is possible to spin up services with greater precision and efficiency. This type of scaling will be critical as 5G services become more dynamic and delivered on-demand.

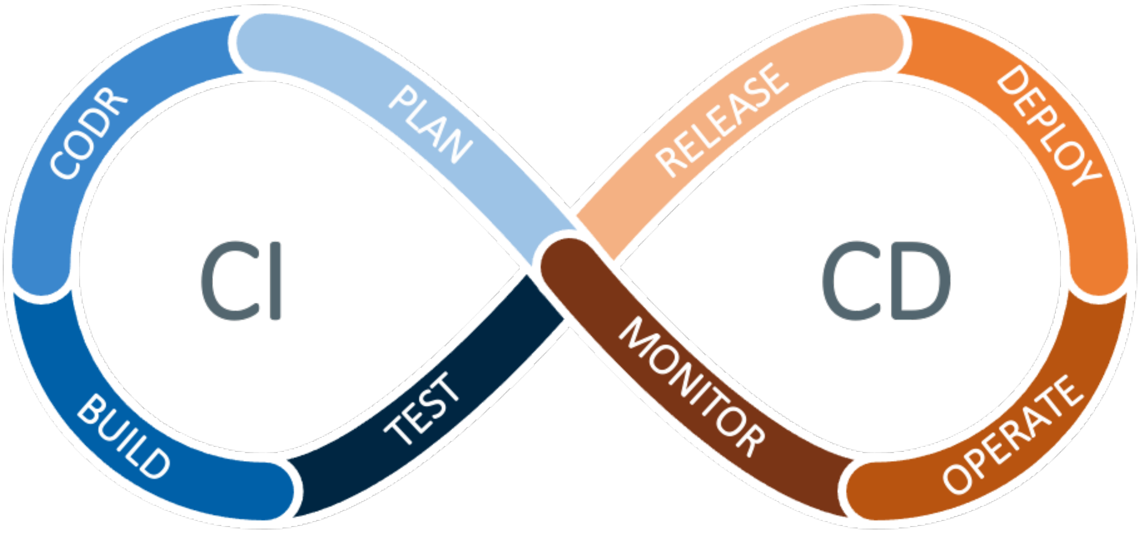


Figure 9 - Continuous assurance development/deployment for 5G will be critical

Also, as a containerized solution Continuous Integration/ Continuous Deployment (CI/CD) processes are used to develop and evolve the product. This means that any new development can be integrated and deployed in a matter of seconds, enabling quick improvements and updates to the probing solution. Moving to cloud-based software development will allow operators to continually adapt their probe layer to keep in line with the advancement of their 5G network and the services running over it.

By deploying containerized probes, operators can deploy and integrate probes into their cloud-native network. Using Kubernetes (K8s) can automatically control the containerized components lifecycle starting from the initial day-0 instantiation and scaling in or out to improve network resource utilization and throughout the platform lifecycle, providing a probing solution that is dynamic and offers the operational agility needed for advanced 5G services.

When probing is deployed as software-controlled functions, they seamlessly integrate into the operators' network as a built-in solution for a cloud-native architecture implemented on public, private, or hybrid clouds. Designed to monitor and troubleshoot CNFs and provide data streams for analytics, hosted in a container, and deployed as close as possible to the monitored traffic (sidecar deployment). Allowing efficient data movement and processing, as data capture and filtering occurs at the source NF, resulting in significant traffic reduction within the infrastructure.

Distributed from the core to the Edge

The 5G core can be deployed centrally and at the Edge. The ability to distribute functions will be necessary to managing traffic growth and to enabling low-latency 5G services that must be hosted close to the user. In 4G EPC, Control and User Plane Separation (CUPS) was introduced to allow for independent scaling, and the same concept is native to the 5G system and core network.

One impact of CUPS is distributing the user plane at the edge data center while the control-plane functions remain centralized. This allows for GTP traffic from the RAN to be terminated at the Edge and then be routed according to the service type. In some cases, the application itself will reside at the same edge cloud location, removing the need for backhaul to the central data center and enabling low-latency services.

Edge deployment represents a significant change to the mobile network architecture. Failure or downtime will be catastrophic and could extend network-wide, resulting in unhappy customers, lost revenue, breached service level agreements (SLAs), and lasting brand damage. Distributed core networks must be monitored by lightweight, containerized microprobes to continually provide real-time subscriber analysis so network degradation can be fixed as soon as possible.

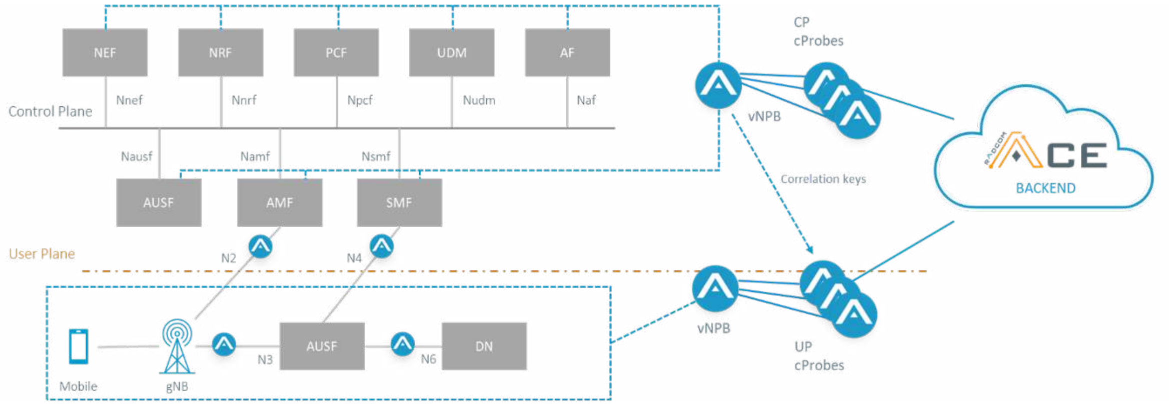


Figure 10 - RADCOM ACE deploys microprobes at the Edge

These microprobes will need to be deployed on demand by the orchestrator. This is important because, in many cases, the edge data center will support multiple services types and may allocate resources dynamically. For example, a given amount of computing and storage capacity may need to be shared dynamically, with each change inducing a corresponding change in the monitoring requirements. Similarly, if a new service is deployed or a new customer-type is on-boarded, the related monitoring capabilities will be required.

Vendor-agnostic, independent auditors

Probes are network vendor-agnostic and show a measurement that is consistent across all the network elements. Providing an independent auditor in that will paint a vendor-agnostic picture of the end-to-end service quality, challenging if relying on network element counters that depend on the network equipment to monitor itself. The data output varies from to vendor.

By deploying a probe-layer as part of the service assurance solution, the operator gains an end-to-end view of the service that prioritizes network issues according to their service and customer impact. By understanding the impact on the subscriber, operators can troubleshoot issues faster and improve customer experience.

Service assurance probes provide operators with an end-to-end view across different network tiers (Edge, core, IP), cross-domains (physical and virtual). They are essential for assuring the customer experience in today’s competitive market, watching all the traffic that flows through the network, and correlating the data into individual customer sessions to understand the service quality experienced in each call and data session (VoLTE call, web browsing, video streaming).

Probing assurance use cases

Mobile Broadband

With the initial rollout of NSA 5G Enhanced Mobile Broadband (eMBB) is one of the primary services being rolled out to both consumers and businesses alike and will drive 5G development and facilitate its success. Operators that deploy NSA dual connectivity to deliver enhanced mobile broadband must support both radios and their interactions. In NSA, scheduling and handovers are steered using the 4G control channel, so for handoffs between RATs, operators must monitor the control plane. 4G devices already have reliable handoff mechanisms, but switching a session from one RAT to another can still have significant latency that might cause sessions to drop during handoff. So, it is critical operators monitor these handoffs to ensure they do not impact the customer experience.

The ability to detect and report on dual connectivity NR capable devices through probing enables the operator to differentiate between the 5G and 4G connected traffic, which is critical when rolling out NSA 5G services. New rules and enrichments separate and correlate both control plane and user plane information for 5G active UEs. By deploying probes, operators can monitor the following KPIs;

- Accessibility, Mobility & Retainability based on S1-MME
- RAT Type field populated with E-UTRAN or NR to filter and group all GTP-U /GTP-C KPIs by 4G/5G access
- By cell distribution (ECGI or NCGI)
- Statistics and ratios of the number of active subscribers, types of devices and their capabilities, modes of operation of those subscribers (4G active versus 5G capable/connected UEs)

Operators can also drill down from these KPIs to a detailed end-to-end correlated call trace for root-cause analysis.

Ensuring end-to-end VoNR and Vo5G service quality

Evolving from VoLTE, VoNR (Voice over 5G New Radio) is the IMS based voice services which use 5G as the access network (as opposed to LTE and VoLTE). As the technology advances, Voice over 5G (Vo5G) services will build on this as evolved voice systems leverage combined 5G core network elements along with IP Multimedia Systems (IMS), VoLTE enhancements, 5G Evolved Packet Core (EPC), and other 5G New Radio (5GNR) 5G radio access network equipment such as smart antennas.

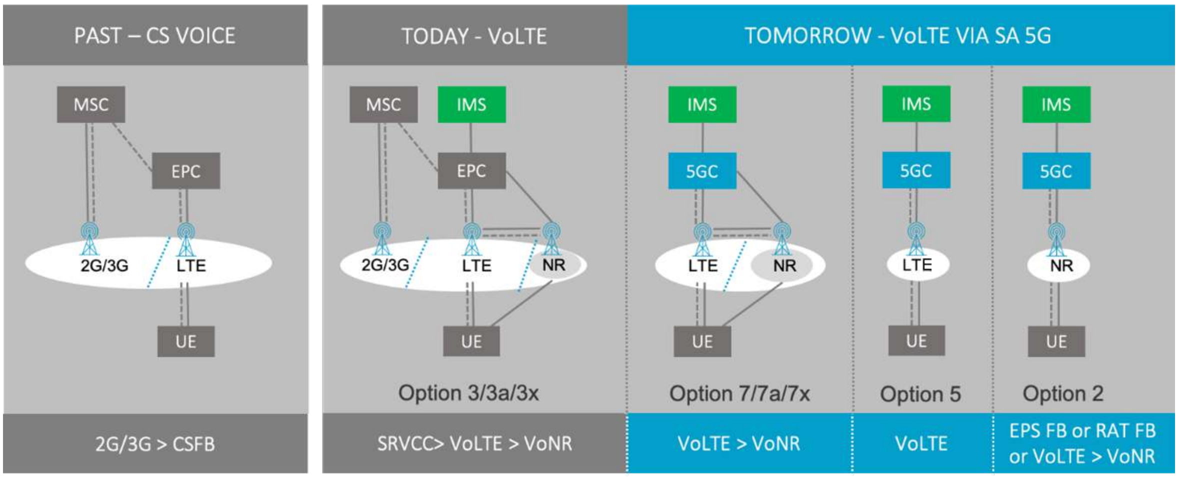


Figure 11 – The evolution of voice services to Vo5G

Advantages of Vo5G services will include ultra-high definition voice/audio for both voice-only calls and integration with applications and content such as announcements, music, conferencing, and more. Vo5G will also provide enhanced support for real-time communications, including Rich Communications Services (RCS) integration. Real-time feature/functionality will include unprecedented interactive capabilities such as real-time language translations. Many of the more advanced functions will work only in a 5GNR environment with 5G core infrastructure support.

Probe-based assurance will be critical for monitoring VoNR to perform end-to-end cross-domain correlation and root cause analysis across RAN, backhaul, core, IMS, and the VoNR application server. Monitoring handovers from Voice over NR (VoNR) to VoLTE when crossing into service areas lacking sufficient 5GNR coverage for a smooth transition will be critical and necessary for isolating network performance issues that cause QoS degradation and avoiding call drops. Degradations can occur at any point in the network. Hence, an operator needs an intelligent end-to-end monitoring system to provide the all-important full network visibility, which proactively highlights when degradations occur across the entire network.

Using probe-data operators to have a complete view of the network will pinpoint which areas require more bandwidth. Suppose there is a heavy concentration of traffic on one cell. In that case, a smart service assurance solution will detect this and trigger an alarm alerting the operator to consider optimizing the network, catching the issue before it becomes a problem.

If operators want to deliver the highest QoE to their customers, they must ensure that their VoNR services are running smoothly. In a cloud-native environment, the only way to truly ensure this is with a probe-based service assurance solution that provides the operator with an end-to-end view of the service quality and pinpoints where degradations are occurring.

Network slicing

Network slicing is a critical use case for 5G and is predicted to be one of the main revenue generators for telecom operators. Therefore, it will be vital for operators to monitor the network effectively and ensure each service is optimized. It will enable operators to offer customers a range of virtual services via the same core infrastructure. This means an operator could use one slice of the network to supply super high-speed broadband for gamers on one slice and low-latency, mission-critical services on a different slice. Both slices will be reliant on the same underlying network.

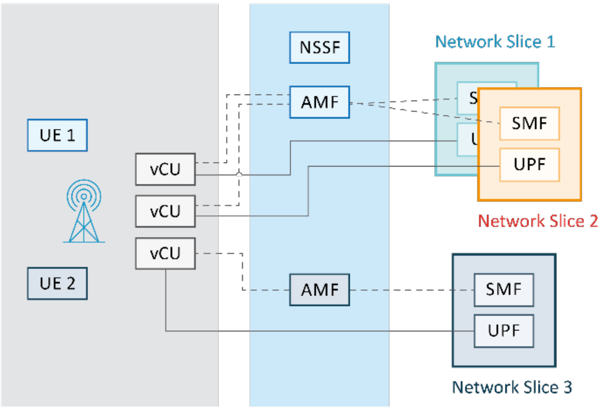


Figure 12 - Deploying virtual services slices on the 5GC

Effectively monitoring the network is essential if operators want to meet the Service Level Agreement (SLA) for each slice, and for a slice that would service-connected cars, for example, could be a matter of life and death. Each slice would have different needs, and so a comprehensive real-time analytics system is crucial for ensuring those needs are being delivered.

The orchestrator composes the network service by selecting the required network functions and configuring them according to the use-case requirements. Services are defined in a service catalog, managed by the service orchestrator, and translated into data models and policies by the network orchestrator. The network orchestrator instantiates the slice from the available resources and instantiates an assurance probe that monitors the service quality.

An integrated service assurance solution will identify the network slice instance and create a slice utilization KPI per network slice instance. An NF consumer, such as the Policy Control Function (PCF) or Network Slice Selection Function (NSSF), can subscribe to or unsubscribe from real-time or periodic notifications KPIs and receive notifications when a KPIs exceed a specified threshold. The PCF takes these inputs from the assurance solution to navigate traffic policies and assign more resources when necessary, enabling the operator to manage the slices dynamically.

In a cloud-native network, services are supported on shared resources. This adds an essential dimension to service assurance in the 5G core relative to a classic EPC deployment for 4G. Unless the resource is significantly over-provisioned at multiple locations (an undesirable scenario), there may be times when different services compete for resources. In this case, the service assurance solution must monitor the slice's SLA and inform the network orchestrator when thresholds are about to be breached.

Network policy can then determine which services have priority and should be moved to a new location or temporarily downgraded. In effect, this means service assurance needs to monitor VNFs and inter-VNF traffic and interwork with the tools monitoring the NFVI cloud platform.
Creating and managing many diverse services composed of many microservices, dynamically running on cloud infrastructure, will take mobile operators into an unfamiliar world, relative to the static world configurations that characterize many core networks today. The sheer number of network events, changes, and options, and their variance over time, will overwhelm human comprehension and their ability to make optimal network policy decisions manually. Automated service assurance, fed by probes, will be required to take the network slicing vision to reality and requires the ability to dynamically deploy probes to monitor the service quality and ensure SLAs are being met.

Integrating probes into the 5G ecosystem

Automation

The sheer number of events and traffic in a 5G network will overwhelm human understanding and their ability to make optimal network policy decisions manually. Automation, rather than human intervention, is required to take the 5G service vision into reality, dependent on a cloud-native approach to service assurance that seamlessly integrates into the 5G core and provides feedback to the operators' orchestration. The ability to assure these diverse service types on a cloud-native platform is fundamental to these services' success. 5G introduces many new interfaces, protocols, and technologies into the cloud core, and dynamic, cloud-native service assurance will be used to support automated, closed-loop service optimization. A closed-loop approach will enable operators to meet SLAs across these multiple service types deliver end-to-end services that meet a unified network policy.

For decades, assurance systems were mainly deployed for network validation and fault and performance issue reportage, leaving the expensive and often manual task of performing root-cause analysis and issue resolution to engineers in the network and operations departments. In its new incarnation for 5G, assurance systems will take on the operational 'nervous system' responsible for driving the network automation and lifecycle management of the 5G services.

Artificial intelligence and machine learning

Part of network automation utilizes Artificial Intelligence (AI) and Machine Learning (ML) to sift through the vast amount of information using machines to provide automated insights. AI has natural advantages over humans in analyzing massive amounts of data and finding patterns and relationships in the data. Machines can:

- Handle repetitive assignments
- Process complicated, multi-dimension tasks
- Process and correlate information from many different sources
- Do not require manual adaptation
- Accumulate experience over time
- Work non-stop 24/7/365

This frees engineers up to spend more time on the critical task of optimizing the network performance and solving network degradations, rather than wasting time looking for issues—machines and humans working together to ensure superior customer experience and operational excellence.

For use cases like dynamic network slicing and predictive analytics that identify traffic patterns and trends to create rules and policies that proactively prevent and resolve issues will be vital in 5G. These insights provided by AI-driven service assurance will feed into the operators' orchestration to enable open/closed-loop control, which saves on OPEX and ensures the network quality automatically, which will be vital to delivering high quality, personalized services in the 5G era.

So, operators must deploy probe-based assurance with built-in modular AI that can apply AI/ML to the data it already collects to provide KPI anomaly detection. Having built-in AI and anomaly detection offers several benefits to operators;

- Data already collected for assurance is used
- AI is applied to all data collected and not a subset
- Saves unnecessary expenses for an additional solution (such as storage costs)
- Saves time massaging the data for external processing
- Runs on any data set (for example, first throughput and in an instant change to the release cause)

With a modular, cloud-native assurance architecture, new and updated ML models can be added to the solution seamlessly as different models will be used for different anomalies. Best-in-breed algorithms and ML models such as Prophet (built and open-sourced by Facebook) can be used for KPI anomaly detection like Release Cause Distribution. In contrast, other ML models are better suited to other use cases (this can include the operators' models). An assurance solution must integrate these different ML models and provide tools in which different ML model results can be compared. The user can utilize other models for different use cases.

Network orchestration

The traditional assurance approach has been mostly unidirectional – that is, process the network events and present the fault and performance data for visualization via dashboards and reports. Following this, operations personnel would analyze the outputs, manually execute a workflow of steps to identify the root causes of the performance and service degradation, and manually carry out the actions to rectify the network issues through configuration changes. Closed-loop automation seamlessly integrates the two sets of processes by triggering policy-driven network changes through the MANO systems such as NFV orchestration, SDN control, and multi-domain WAN configuration. This type of closed-loop automation will be critical for managing 5G networks and will require probe-based assurance to feed automation mechanisms controlled by the management and orchestration (MANO) layer. In this way, service assurance becomes critical to the real-time operation of the 5G core.

A cloud-native assurance solution is integrated into orchestrators such as KVM hypervisor, OpenStack VMWare, and Kubernetes (K8s) to control the assurance solution's components lifecycle starting from the initial day-0 instantiation and throughout the overall platform lifecycle. This integration between assurance and the orchestrator enables assurance automation (from instantiation, healing, upgrades, modification, and scaling) and end-to-end service quality management. Assurance stays in tune with the ebb and flow of the network.

When new services and network functions are launched or grow, assurance can be automatically instantiated and scaled in/out to respond to changes in prevailing network conditions and ensure new additions to the network are automatically monitored, and cloud resources managed efficiently. From the 5G Core to the Edge the assurance solution will need to be automated to instantiate and scale across multiple MEC/Edge user plane sites and central control plane cloud sites.

Assurance also needs to utilize auto-instantiation mechanisms, auto-scaling, and re-balancing according to required trace policies and to 5G NF service discovery. This assurance automation includes auto-adjustment to the monitoring trace policy to fit available monitoring resources based on the system health and traffic monitoring KPIs. Monitoring and logging of all system components should be performed for interfacing to the orchestrator to report system health. This will also lead to a closed-loop solution for system availability so that system health issues can be resolved automatically, similar to the automatic scaling process.

Northbound Mediation

Probe-based assurance solutions also provide additional methods and formats for data export that ensures seamless and flexible integration with OSS/BSS solutions. Examples of this type of integration's benefits are to trigger a CTTS ticket when more than five-speed test results are under the baseline threshold or to publish a CTTS ticket event topic to a CTTS Kafka consumer.

For northbound mediation, assurance solutions can utilize Apache NiFi that (to automate the flow of data between different software systems) to:

- Deliver alarms/events via streaming analytics:
    - Publish to a queue (Kafka, JMS, AMQP)
    - Send an email
    - Call a Rest API
    - Report to Syslog

- Deliver periodic data for upstreaming processing
    - Write to an SQL or HBase database (as described above)
    - Write to file (CSV, JSON)
    - Additional reporting via the BI-layer

Real-time Streaming Analytics

Automated assurance for 5G also needs to provide customizable real-time streaming analytics for network intelligence fed to the operators' orchestration to detect, analyze, and resolve issues automatically. Real-time stream processing consumes and publishes Kafka topics and can combine this data with external data sources such as:

- KPIs generated by service-assurance probes
- Events from incident and change management systems
- Statistics

Streaming analytics will be critical for 5G and can be used by operators for;

- Time series aggregation (e.g., sliding windows)
- Time triggers for data export
- Rules (discard or save records by baseline thresholds)

The ability to process and analyze streaming data dramatically improves time to insight—which will be a critical component for automating 5G networks and assuring 5G service quality. Some examples of how real-time steaming will be used in 5G;

- Network slicing
  The auto-scaling capability of the streaming platform can be leveraged for data-type variability and volume variability associated with various network slices, ensuring QoS for each slice

- Network operations
  This will need to be data-driven to achieve maximum automation. This means capturing, processing, and reacting to network data in real-time. Streaming analytics can also help establish a baseline of network performance, traffic flows, and user mobility, responding to both gradual changes to the baseline and anomalies that enable predictive operations.

RADCOM ACE

RADCOM ACE is a powerful combination of cloud-native and containerized Service Assurance and AI-driven Network Insights. Together they enable telecom operators to gain full network visibility across multiple network domains (2G, 3G, 4G, and 5G) and cloud environments (public, private, and hybrid). These products combine to provide the operator with an understanding of what is happening in their network 24/7. While at the same time minimizing network operational and capital expenses, delivering low-level tools to help resolve network degradations, and offering an end-to-end view of the customer experience across all their services, assisting operators to ensure superior customer experience and optimize service quality.

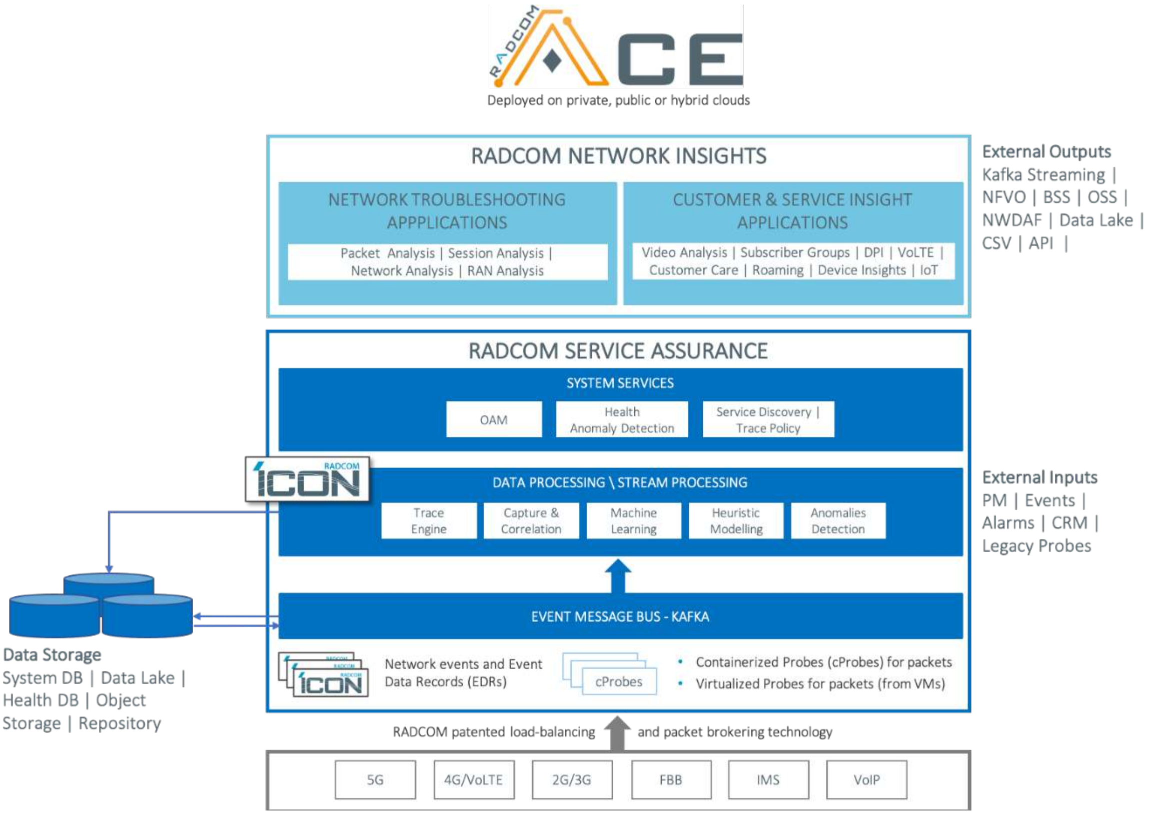

RADCOM ACE
Deployed on private, public or hybrid clouds
RADCOM NETWORK INSIGHTS
NETWORK TROUBLESHOOTING APPPLICATIONS
Packet Analysis | Session Analysis | Network Analysis | RAN Analysis
CUSTOMER & SERVICE INSIGHT APPLICATIONS
Video Analysis | Subscriber Groups | DPI | VoLTE | Customer Care | Roaming | Device Insights | IoT
External Outputs
Kafka Streaming | NFVO | BSS | OSS | NWDAF | Data Lake | CSV | API |
RADCOM SERVICE ASSURANCE
SYSTEM SERVICES
OAM
Health Anomaly Detection
Service Discovery | Trace Policy
RADCOM ICON
DATA PROCESSING \ STREAM PROCESSING
Trace Engine
Capture & Correlation
Machine Learning
Heuristic Modelling
Anomalies Detection
External Inputs
PM | Events | Alarms | CRM | Legacy Probes
EVENT MESSAGE BUS - KAFKA
Data Storage
System DB | Data Lake | Health DB | Object Storage | Repository
ICON
Network events and Event Data Records (EDRs)
cProbes
Containerized Probes (cProbes) for packets
Virtualized Probes for packets (from VMs)
RADCOM patented load-balancing and packet brokering technology
5G
4G/VoLTE
2G/3G
FBB
IMS
VoIP

Figure 13 - RADCOM ACE Solution Architecture

RADCOM ACE is an entirely containerized, cloud-native portfolio and, therefore, future-proof and fully supports 5G (standalone and non-standalone). RADCOM's robust and agile portfolio is built using a microservices architecture allowing for scaling, updating, or even the complete replacement of each part and offers real-time performance and elastic scalability. RADCOM works with leading orchestration platforms (such as Kubernetes) and is part of partnership programs such as Cisco, Amdocs Network Cloud Service, Huawei Fusionsphere, Intel Network Builders Program, ONAP, OpenSource Mano (OSM), Nokia Cloudband Ecosystem, Telefonica UNICA, and VMWare Technology Alliance Partner.

RADCOM ACE advantages;

- AI-driven Insights Use of AI, machine learning, and heuristic modeling to proactively and predictively monitor and troubleshoot the network with automated root-cause analysis, anomaly detection, and insights into encrypted traffic for such services as video streaming
- Automated Assurance Automates solution deployment for on-demand instantiation, scaling, healing, and updating for a closed-loop approach to assurance with Kubernetes controlling the containerized components lifecycle
- Built for 5G Full support for SA 5G (auto-service detect, SBA architecture, SBI deciphering and complex CUPS correlation, Packet Forwarding Control Protocol (PFCP), new aggregations, and dimensioning)
- Cloud-Native Architecture An entirely cloud-native, containerized portfolio, built using a microservices-based architecture allowing for scaling, updating, or even the complete replacement of each part and offers high real-time performance, elastic scalability, and resilience with stateless and lightweight functions
- Real-time Streaming Analytics Delivers automated real-time network intelligence streamed to the operators' orchestration to automatically detect, analyze, and resolve issues.

RADCOM's solution is deployed at multiple operators globally, such as AT&T, Beeline, Globe, Rakuten, and Telefonica and has received wide industry recognition; winning a Frost & Sullivan Product Differentiation Innovation Awards three times, winning multiple TMC Labs Innovation Awards, and winning the TMC Award for NFV Innovation. RADCOM ACE is the most advanced service assurance solution in the market, deployed as a multiple Cloud-Native Functions (CNFs), and automated so the solution can be instantiated in minutes with low-touch deployments and on-demand probing, all essential in a 5G environment. RADCOM utilizes advanced, cutting-edge technology such as AI and machine learning to monitor and troubleshoot for network anomalies proactively, automatically perform root cause analysis, and provide insights into encrypted traffic for such services streaming video services, tethering, and gaming.

RADCOM's solution enables operators to effectively handle significant expansions and technology transformations while replacing legacy probe solutions. RADCOM is open to discussing legacy swap-outs (if required). RADCOM also offers a unique cloud-native business model, based on functionality and technology and not traffic or subscriber numbers, marking an end to linear pricing. The model is simple and based on a multi-year annual subscription fee. Furthermore, RADCOM ACE significantly minimizes the operational cost and complexity involved with monitoring an operator's network. The solution provides a versatile solution with efficient resource usage, ranging from a centralized deployment with minimal resources to a full-scale distributed solution.

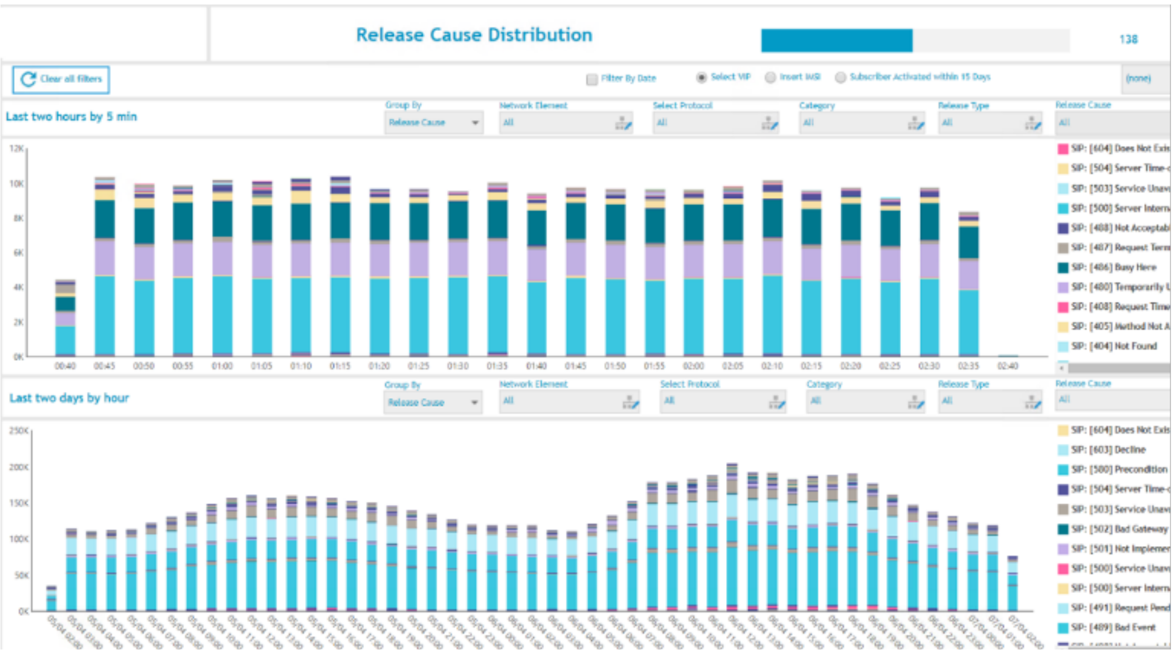
Built-in AI for anomaly detection

This section provides an example of RADCOM's built-in AI for anomaly detection using ML to provide network alerts based on analysis of Release Causes across multiple network elements. It is currently deployed at RADCOM customer sites to alert on potential issues in near real-time.

Release Cause Anomaly Analysis

Partial network outages result in a loss of service and a drop in perceived customer experience for many subscribers. It is imperative to identify disruptions in real-time and provide tools to isolate and analyze the outage's cause and apply corrective measures before customer experience is impacted.

Network outages are associated with a relative increase in the release cause count reported by the core network's various network elements. RADCOM's system continually monitors the release cause count between all network elements and applies advanced machine learning algorithms to determine whether an outage anomaly has occurred. The number of impacted subscribers determines the severity of this anomaly, and an alarm is triggered accordingly. Once this alarm has been received, the network engineer is directed to the Release Cause Distribution dashboard to analyze the outage's cause and apply corrective measures.



Release Cause Distribution

138

Clear all filters

☐ Filter By Date ● Select VIP ○ Insert IMSI ○ Subscriber Activated within 15 Days (none)

Last two hours by 5 min

Group By: Release Cause | Network Element: All | Select Protocol: All | Category: All | Release Type: All | Release Cause: All

Last two days by hour

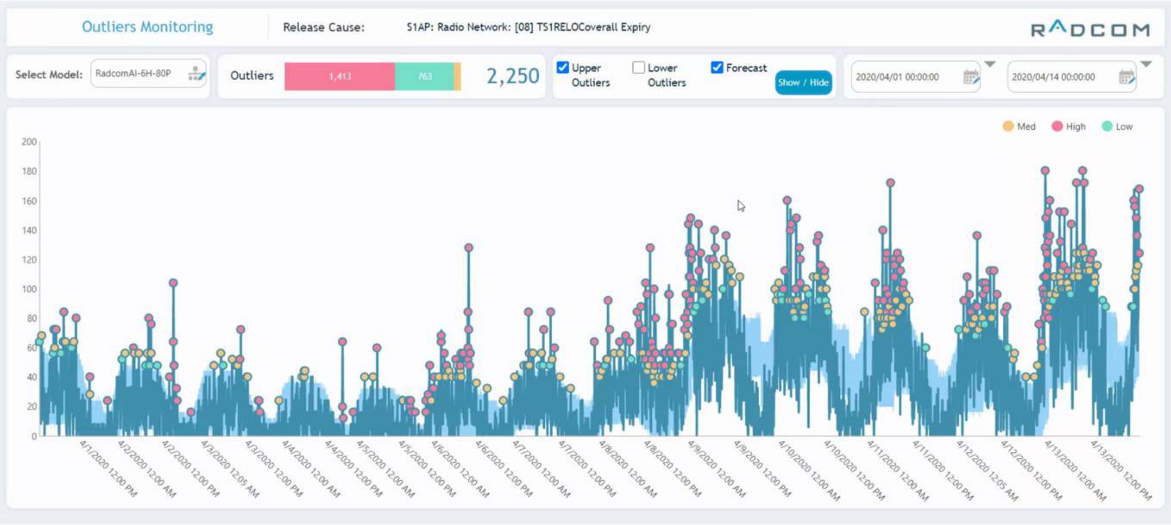
Group By: Release Cause | Network Element: All | Select Protocol: All | Category: All | Release Type: All | Release Cause: All

Figure 14 - Release Cause Distribution Dashboard

Release Cause Distribution Dashboard

RADCOM's Release Cause Distribution dashboard provides a nearly real-time view of the release cause count between core network elements for all major protocols, including S1AP, Diameter, SGs, GTPV1, and two and VoLTE SIP.

The upper pane is refreshed every 15 minutes and shows the current release cause distribution for the select network elements and protocol. The same dashboard may also be filtered on a specific subscriber. The number of impacted subscribers is provided for a selected release cause to view the list of subscribers. The user may also drill to view the signaling messages between the source and destination network elements.



Outliers Monitoring

Release Cause: S1AP: Radio Network: [08] TS1RELOCoverall Expiry

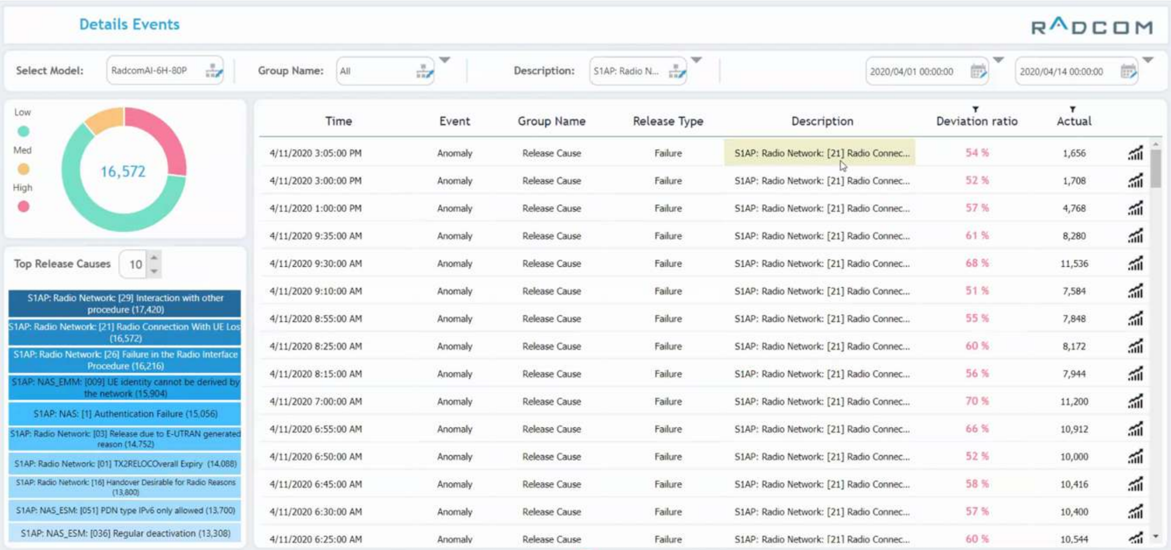
Select Model: RadcomAI-6H-80P

Outliers 1,413 763 2,250

☑ Upper Outliers ☐ Lower Outliers ☑ Forecast

2020/04/01 00:00:00 2020/04/14 00:00:00

Figure 15 - Viewing the baseline, lower/upper boundaries, and the detected anomalies



Details Events

Select Model: RadcomAI-6H-80P

Group Name: All

Description: S1AP: Radio N...

2020/04/01 00:00:00 2020/04/14 00:00:00

Top Release Causes 10

- S1AP: Radio Network: [29] Interaction with other procedure (17,420)
- S1AP: Radio Network: [21] Radio Connection With UE Los (16,572)
- S1AP: Radio Network: [26] Failure in the Radio Interface Procedure (16,216)
- S1AP: NAS_EMM: [009] UE identity cannot be derived by the network (15,904)
- S1AP: NAS: [1] Authentication Failure (15,056)
- S1AP: Radio Network: [03] Release due to E-UTRAN generated reason (14,752)
- S1AP: Radio Network: [01] TX2RELOCOverall Expiry (14,088)
- S1AP: Radio Network: [16] Handover Desirable for Radio Reasons (13,800)
- S1AP: NAS_ESM: [051] PDN type IPv6 only allowed (13,700)
- S1AP: NAS_ESM: [036] Regular deactivation (13,308)

Time	Event	Group Name	Release Type	Description	Deviation ratio	Actual
4/11/2020 3:05:00 PM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	54 %	1,656
4/11/2020 3:00:00 PM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	52 %	1,708
4/11/2020 1:00:00 PM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	57 %	4,768
4/11/2020 9:35:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	61 %	8,280
4/11/2020 9:30:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	68 %	11,536
4/11/2020 9:10:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	51 %	7,584
4/11/2020 8:55:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	55 %	7,848
4/11/2020 8:25:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	60 %	8,172
4/11/2020 8:15:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	56 %	7,944
4/11/2020 7:00:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	70 %	11,200
4/11/2020 6:55:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	66 %	10,912
4/11/2020 6:50:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	52 %	10,000
4/11/2020 6:45:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	58 %	10,416
4/11/2020 6:30:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	57 %	10,400
4/11/2020 6:25:00 AM	Anomaly	Release Cause	Failure	S1AP: Radio Network: [21] Radio Connec...	60 %	10,544

Figure 16 - Viewing details of the KPI network anomalies

ML-based Anomaly Detection and Alerts

An advanced algorithm is applied to identify anomalies in the release cause count between all network elements and the associated severity. The anomaly severity may then trigger alarms. This engine can remove ‘outliers’ from the release cause baseline utilizing sophisticated algorithms. The baseline outlier removal facilitates an accurate baseline prediction, improves the detection of network anomalies, and removes ‘false positives.’ We can see the baseline lower/upper boundaries and the detected irregularities in the following forecast, which may be found above. The unique RADCOM algorithm effectively reduces false positives and divides anomalies into three severity levels based on the baseline deviation.

Summary

RADCOM is the leading expert in cloud-native, containerized, and automated probe-based service assurance with AI-driven insights providing operators with an end-to-end view of the service quality and real-life customer experience. RADCOM’s solution supports operators in their transition to 5G by delivering dynamic, on-demand service assurance and network troubleshooting at a macro and micro level so customer-affecting network degradations can be resolved quickly with minimal effort. RADCOM’s solution for 5G, RADCOM ACE, is explicitly designed for telecom operators and delivers Automated, Containerized, and End-to-end network insights. The solution seamlessly integrates into an operators’ cloud infrastructure. It is controlled by leading network orchestration solutions such as Kubernetes (K8) that manage all the assurance microservice components' lifecycle for a closed-loop and automated approach to service assurance.

For more information about RADCOM ACE, visit: https://www.radcom.com/radcom-ace