RADCOM ACE

AI-driven assurance for ensuring 5G service quality

Publication Date: November 2020

Website: http://www.radcom.com

Table of Contents

Introduction

5G will usher in a more wireless world where we are always connected, and services are available on-demand. We will be able to do more on the go, and tasks that would take minutes on a 4G network, will take seconds on 5G. Blistering download, upload, and streaming speeds will enhance the feeling of being always connected. Services such as 3D video augmented and virtual reality, massive multi-player cloud-gaming, smart industry automation, and self-driving cars. Mission-critical applications like critical communications and remote healthcare will enrich our lives, possibly save them as well.

These services will be delivered using a new 5G core built on a cloud-native architecture, with multi-access edge computing (MEC) and network slicing to enable the performance, latency, and quality of service expected from 5G networks and create new revenue streams. Data usage increased exponentially in the 4G era, but the unit price of traffic per bit decreased continuously. In 5G, traditional data services alone will not be enough.

Operators need to manage their networks efficiently while bringing new revenue growth. However, the shift to a 5G network is a seismic transformation of the network architecture and its management. However, this brings many benefits to the operator in operational flexibility and scalability while laying down the foundations for the next generation of exciting, dynamic services. This technology change presents many more complexities and challenges that the operator will need to overcome to make this transformation a success.

Add to this the significant increase in data traffic and millions of more devices connecting to the cloud; operators have substantial challenges in sifting through vast amounts of data to pinpoint service degradations, understanding the end-to-end customer experience, and troubleshooting the network performance.

Looking for a needle

In the beginning, mobile networks were more straight forward. The only service on 2G was voice that was designed and packaged for the subscriber by the telecom operator. Network engineers needed to monitor a handful of different KPIs to ensure the service was running correctly. As various iterations of mobile technology have rolled out, more and more complex layers have been added. Now, hundreds of different services are running over the network, voice services (normally IP-based VoLTE) and data services that cover a wide range of use cases from video streaming to social media platforms. A significant number of these services are no longer designed and packaged by the operator, but Over-The-Top (OTT) services like Netflix or WhatsApp. Even though many services delivered are no longer from the operator, we are in an age where it is easy for

subscribers to churn, and service expectations are high. As 5G begins rolling out, this situation is not going to change anytime soon.

Recent statistics show that between Q2 2019 and Q2 2020, network data traffic grew 53%[1], and this trend is expected to continue. More people will connect (forecast to be 5.3b people by 2023, up from ~4.5 today). More devices will connect (predicted to be 29.3b devices by 2023, up from ~20b today), and all this creates more and more traffic, which is expected to grow by 27% annually between 2020 and 2025. This data traffic will be taken up to a whole new level with 5G. For example, watching a streaming sports event in multi-view could consume about 7GB per hour, while a high-quality AR/VR stream with a media (bit) rate of 25Mbps would consume as much as 12GB per hour.

Network traffic growth is driven by both the rising number of smartphone subscriptions and an increasing average data volume per subscription, fuelled by increased viewing of video content. Monitoring all this data and pinpointing customer-affecting network degradations will be like looking for a needle in a field of haystacks. With so many different services, devices, network traffic, and combinations, traditional service assurance solutions with manual processes will fail in 5G.

70% of operator time is devoted to the discovery and root cause of network issues[2]

With operators already spending significant amounts of time isolating service degradations and troubleshooting the network, how are they expected to manage the transition to 5G and the clear jump in the data load while managing the network more dynamically?

[1] Ericsson Mobility Report - June 2020

[2] Analysys Mason - ML/AI-based automated assurance is critical for the success of 5G

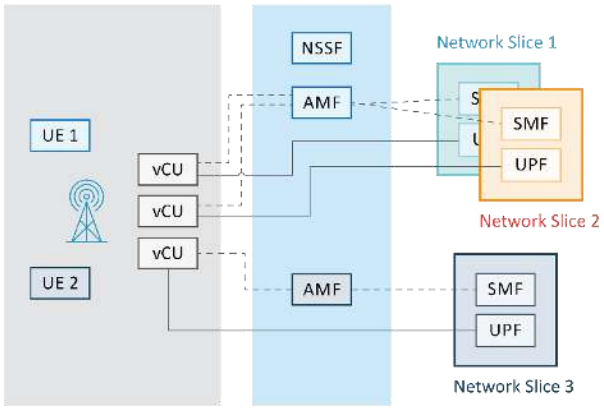


Figure 1 - Advanced 5G services like network slicing will require automation

It is apparent that for 5G, the traditional telecom network operations and management model will not meet the increasing requirements needed to ensure a smooth transition to 5G. The 5G era requires an intelligent, more automated network. This will allow operators to manage their networks more effectively and deal with the increasing complexity and enable innovative business models, like network slicing and the dynamic management of networks, to provide a more personalized customer experience. Introducing intelligence into the network through Artificial Intelligence (AI) is the key that will unlock the ability to offer quality, dynamic services at scale and deliver on the promise of 5G.

AI is the key

Figure 2 - - The development of Artificial Intelligence (AI)

Since the birth of AI in 1956, AI technologies have become increasingly mature, and AI applications have been expanding. (Machine learning or ML is an application of AI that provides systems the ability to learn and improve from experience without being explicitly programmed automatically. Machine learning focuses on accessing data and using it to learn for itself. An example of this is how Google Photos recognizes dogs in your photos).

With the development of fast, low-cost computing and cloud technologies, machines can analyze massive amounts of data generated by networks every day and embed it into next-generation service assurance solutions.

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 needles—machines and humans working together to ensure superior customer experience and operational excellence.

Critical for automation

AI will be essential for network operations and managing the customer experience in 5G. Furthermore, for more advanced 5G services like dynamic network slicing creating rules and policies that proactively prevent and resolve issues will be vital. All this is part and parcel of the operators' goal to enabling an open/closed-loop approach to network management. Automated insights provided by AI will feed into the operators' orchestration to allow this approach, which saves on OPEX and ensures the network quality automatically, which will be required to deliver high quality, personalized services in the 5G era. So, how do operators best acquire these AI-driven insights into their network?

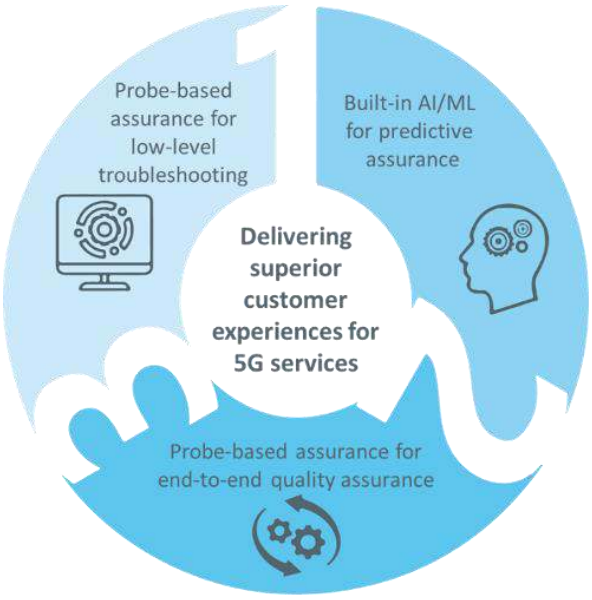


Figure 3 – Three critical assurance pillars for delivering superior 5G service quality

Integrating AI into cloud-native assurance

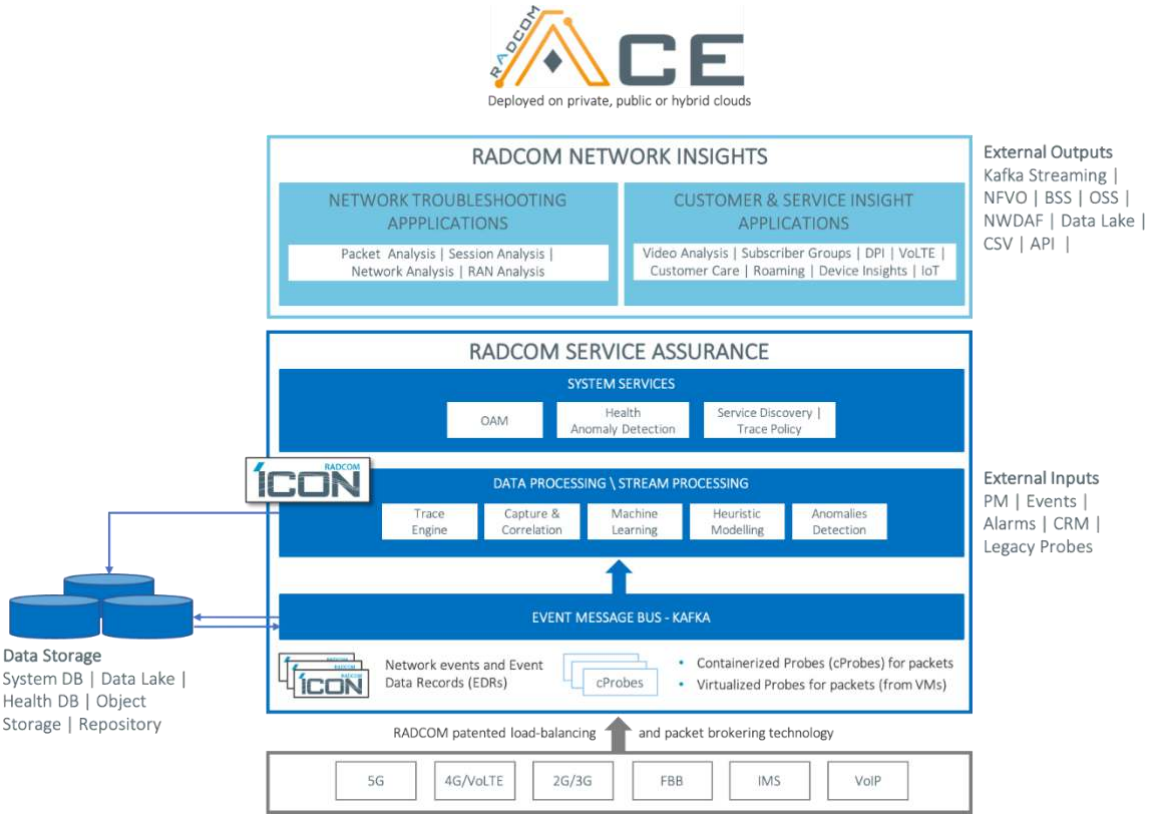
By deploying cloud-native service assurance solutions with built-in AI/ML, the operator can utilize the data already collected through the solutions' containerized probes. AI/ML is then applied as the data is collected rather than deploying an additional solution for AI purposes. Having service assurance with built-in AI 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 instantly change to the release cause)

However, these next-generation assurance solutions need to come with modular architectures that are machine-learning-friendly. New and updated ML models can be seamlessly updated and integrated into the solution. Best-in-breed algorithms and ML models such as Prophet for forecasting time series data (built and open-sourced by Facebook) should be used. Also, the ability to choose from various algorithms and ML models so that the right ones can be used for a different use case is essential. Here are some of the use cases in which operators can use AI-driven assurance to assure their networks automatically.

RADCOM ACE: AI-driven assurance

RADCOM ACE is an automated 5G assurance solution that seamlessly integrates into multiple cloud environments (public, private, and hybrid). It supports advanced 5G assurance capabilities for end-to-end visibility into the customer experience and service quality for 5G. RADCOM ACE also provides operators with AI-driven insights to proactively monitor and troubleshoot the network with KPI anomaly detection.



RADCOM ACE includes the following containerized solutions: RADCOM Service Assurance and AI-driven RADCOM Network Insights. RADCOM Service Assurance captures packet-based data and network events using the following methodologies:

- RADCOM Containerized Probes (cProbes) provide a solution for packet monitoring in 5G and 4G networks.
- Our patent-pending, on-demand solution, RADCOM I.C.O.N., (delivers Intelligent, Container-based, On-demand, Network Analysis from the RAN to the Core) captures and monitors multiple types of network events and Event Data Records (EDRs).

As part of its service-based architecture, RADCOM ACE utilizes an event or service bus that acts as a communication system between the loosely coupled microservices components that can be independently deployed as a distributed system. The Event Bus (Kafka-based technology) has multiple communications purposes:

- Control and management messages
- User activities: trace requests and results
- Data collections: EDRs and packets
- Distributions of microservice states
- Publish/subscribe to internal system events

The Event Bus use enables A-Sync and scaled data collection, processing, and trace request across the distributed system. The architecture supports multiple and parallel publishers and subscribers to the message bus to allow system scalability. The Event bus is assumed to be a persistent store that allows continuous deployment and scaling with zero data loss.

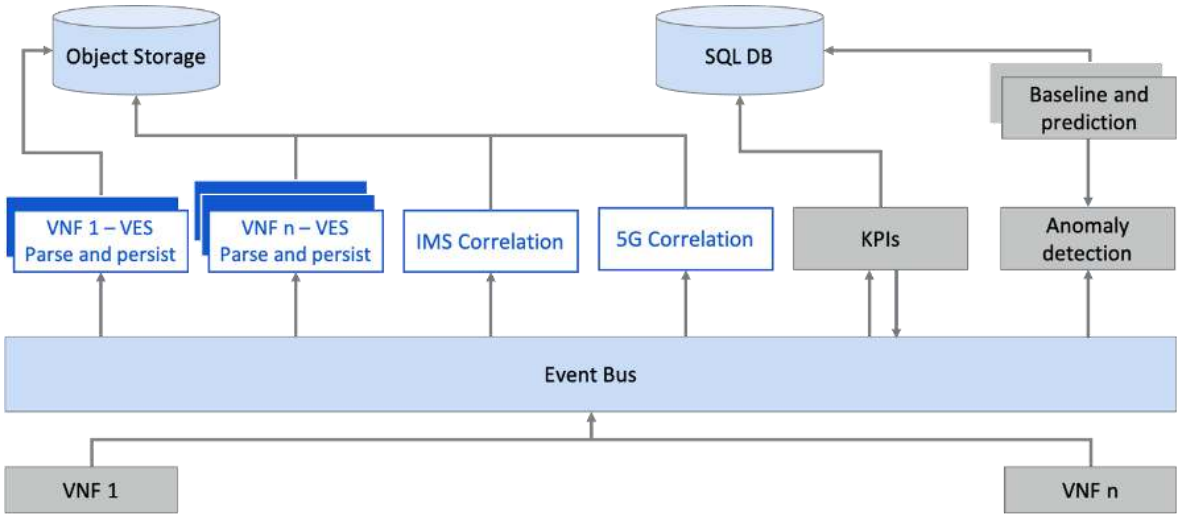


Figure 4 - RADCOM AI is built into the solution

RADCOM ACE uses the event bus to process the anomaly detection in the following workflow:

- The KPI RADCOM I.C.O.N. microservice consumes the EDRs from the event bus and returns the KPI results to the event bus into a dedicated event bus topic for the KPI results
- The baseline and prediction RADCOM I.C.O.N. microservice calculates the predictive value for the requested KPIs using the database and a predictive algorithm. The predictive values are sent to the anomaly detection RADCOM I.C.O.N. microservice
- The anomaly detection the RADCOM I.C.O.N. microservice consumes is the KPIs that results from the event bus, which is then compared to the predictive value, and triggers a series of proactive actions on any outlier such as mail, SNMP, close loop (e.g., change trace policy)

Acing AI-driven assurance

Built-in, modular AI

With built-in modular AI, RADCOM ACE utilizes the data collected through its containerized probes (for network packets) or RADCOM I.C.O.N. (for network events or event data records). It applies AI/ML to offer multiple, telecom specific use cases.

With its modular architecture that is machine learning-friendly, RADCOM ACE can add new and updated ML models to the solution seamlessly. As part of the solution, RADCOM ACE utilizes best-in-breed algorithms and ML models such as Prophet for forecasting time series data that was built and open-sourced by Facebook.

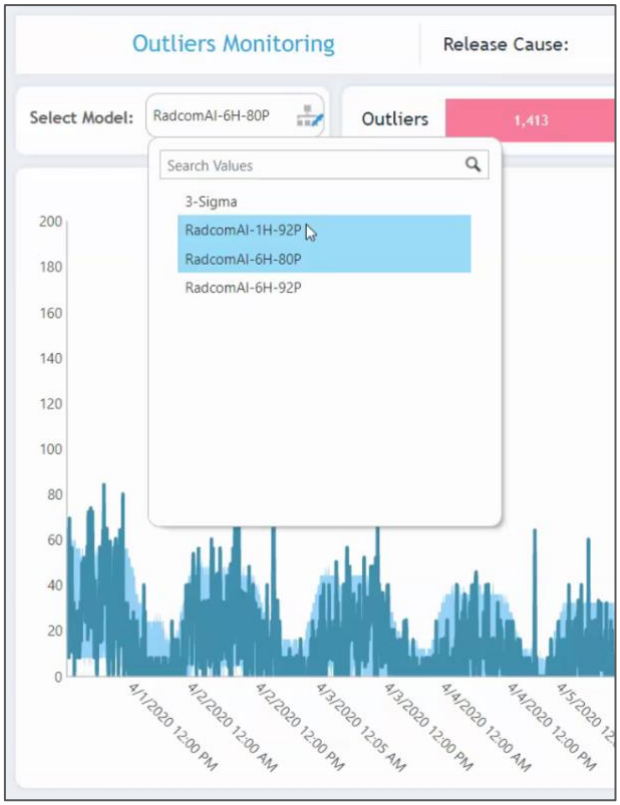


Figure 5 - RADCOM ACE enables operators to choose which ML models to use

These built-in ML models are adapted for telecom-specific use cases and tested. However, the ML models can vary depending on the use case and the users' needs so the telecom operator can integrate their ML models into RADCOM ACE. Also, RADCOM provides tools in which different ML model results can be compared, and the user can utilize other ML models for different use cases.

Implemented in near real-time and batch mode

In the 5G era, it will be critical for operators to be proactive, and RADCOM ACE enables near real-time notifications for use cases such as KPI-based anomaly detection. Usually, operators receive alerts from counters within a range of about 15 to 30 minutes if there are sudden network degradations. Using RADCOM ACE, near real-time notifications can be sent to network engineers for critical issues 5 minutes after a sudden incident. Users can subscribe to notifications (facilitated through the event bus) in real-time. If an important alert happens, different departments within the operators' organization, such as the NOC, can listen and receive these critical alerts and proactively correct network degradations before even the subscribers become aware of them.

RADCOM ACE can also provide the AI-driven insights for use cases like anomaly detection in batch mode so a user can work on a dataset for, say, a week or a month. Then, analyze the data and attain the insights (for example, engineers who do the network dimensioning and planning the network and analyze data over a more extended period.

Filtering rules

RADCOM ACE allows teams to control the amount of data being delivered to them so they don't become overwhelmed and can focus on top priority tasks and are truly customer-affecting according to the human resources available and the time available. For example, an admin user may have ten network engineers available for a given period. They can filter the traffic for different departments (such as engineering or operations) by;

- Type of service
- Severity
    - Critical
    - High
    - Medium

That way, the admin user can decide that the engineers deal with the highest severity issues for the operators' video streaming service. RADCOM provides the operator with default filters, but the operator can adjust these.

RADCOM ACE: AI-driven use cases

This section provides telecom-specific use cases enabled by RADCOM ACE and its AI-driven insights.

Optimize Network Performance

Smartly Plan Network Capacity

Improve the Customer Experience

Optimizing network performance

Anomaly or outlier detection is a method of searching for data that does not match an expected behavior or a pattern in a given data set. Network anomaly detection methods can be classified into five categories:

- Statistical-based
- Classification-based
- Clustering and Outlier-based
- Soft computing-based
- Knowledge-based

Release Cause Anomaly Analysis

An example of KPI-based anomaly detection is for release cause. An advanced algorithm is applied to identify anomalies in the release cause count between all network elements and the associated severity over time. It is thus removing 'outliers' from the release cause baseline. The baseline outlier removal facilitates an accurate baseline prediction, improves the detection of network anomalies, and removes false positives. Also, in calculating a baseline, and a confidence area, it is possible to see points in time beyond the confidence area based on past data. These exceptions can be translated into additional alerts. By utilizing these AI-driven insights, engineers remain focused on handling critical customer-affecting issues rather than handling "alerts" that have no real effect on subscribers.

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.

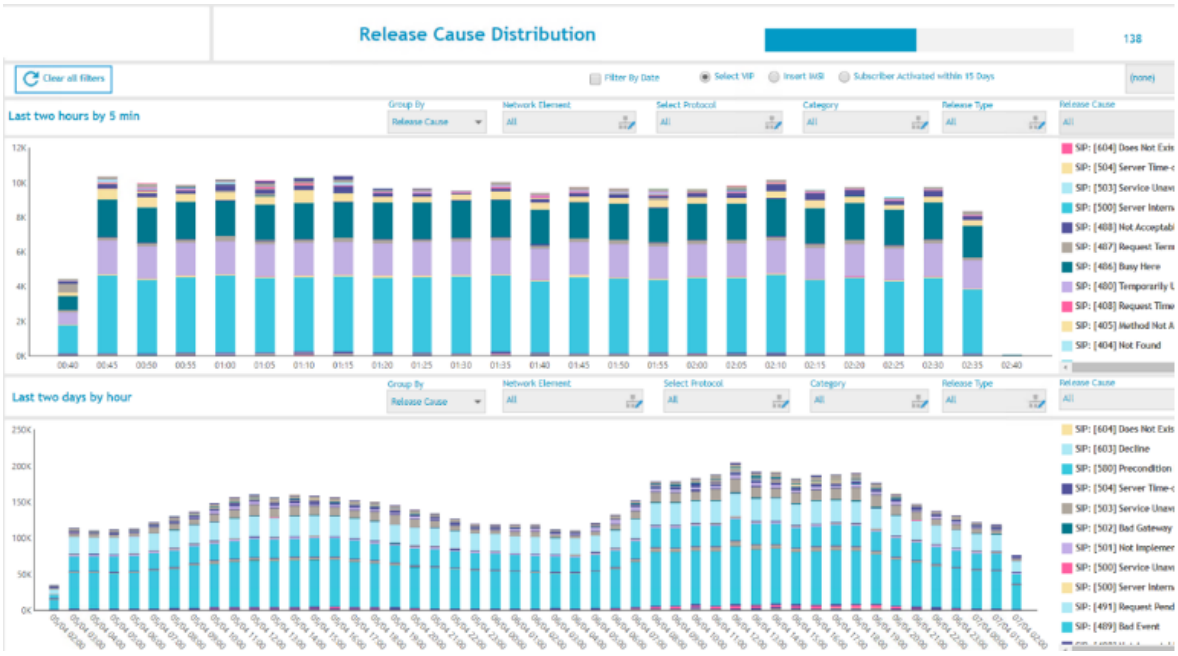


Figure 6 - Release Cause Distribution Dashboard

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 for 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.

Figure 7 - Viewing the baseline, lower/upper boundaries, and the detected anomalies

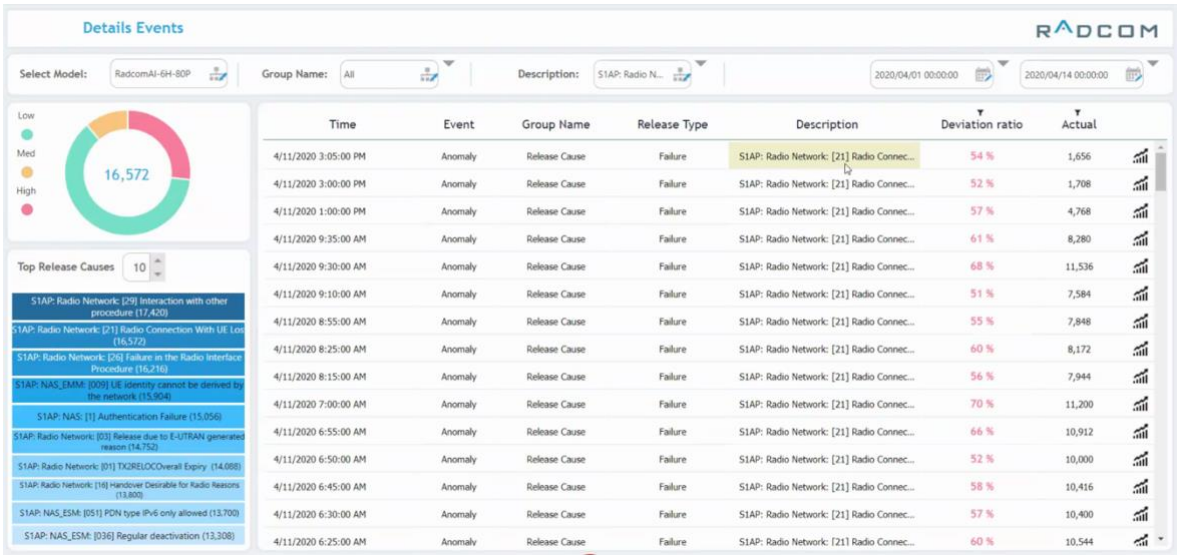


Figure 8 - Viewing details of the KPI network anomalies

Smartly plan network capacity

After generating baselines in different network performance and service quality areas, AI-driven assurance can use this information for more use cases. The most common is predictive analytics. As the solution already knows how to generate the right forecast in the short-term (for say, release cause), it's also possible to create a long-term forecast for various performance and quality indicators and use it to plan future network capacity smartly. By building these processes at the early stages of 5G, engineers will learn to trust the insights that AI can provide and integrate them into network management.

It is allowing more data-driven decision making that will be critical for 5G network management and operations. Furthermore, by incorporating more sophisticated AI algorithms, which consider seasonal seasonality, it will also be possible to create forecasts for an extended period. For example, one month or a whole quarter.

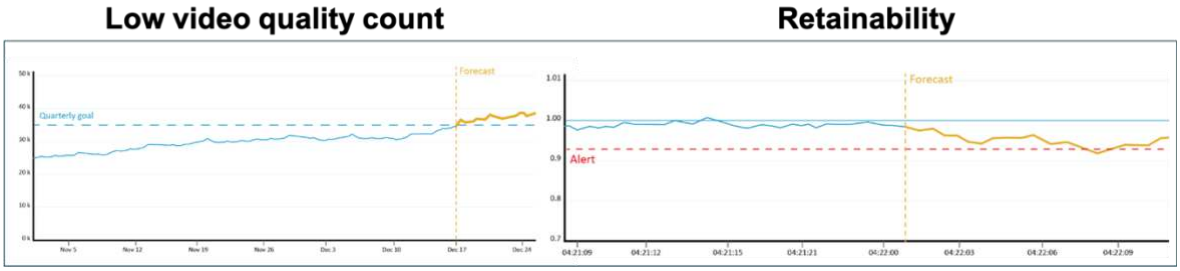


Figure 9 - Predict upcoming network degradation for corrective actions

Improve the customer experience

With more and more encrypted data, operators have challenges in understanding their subscribers' Quality of Experience (QoE). This means that service degradation can occur and go unnoticed by the operator. When this happens (especially with subscribers that heavily use OTT applications available on other networks), subscribers will blame the operator. They could churn even though a 3rd party provides the overall service.

To negotiate this landscape, an operator needs to embrace the trend for privacy and gain accurate, real-time, application-level visibility into what is happening in their network and understand their subscribers' quality of experience.

RADCOM's solution utilizes AI capabilities and cutting-edge ML and heuristic modeling to understand the perceived Quality of Experience (QoE) for multiple use cases:

For video streaming

It is estimated that by 2021, video services will account for 82% of all the traffic on the internet. With 60% of all video data being encrypted and the expectation that this figure will rise, operators face the challenge of extracting insights from the network and understanding the customer's QoE.

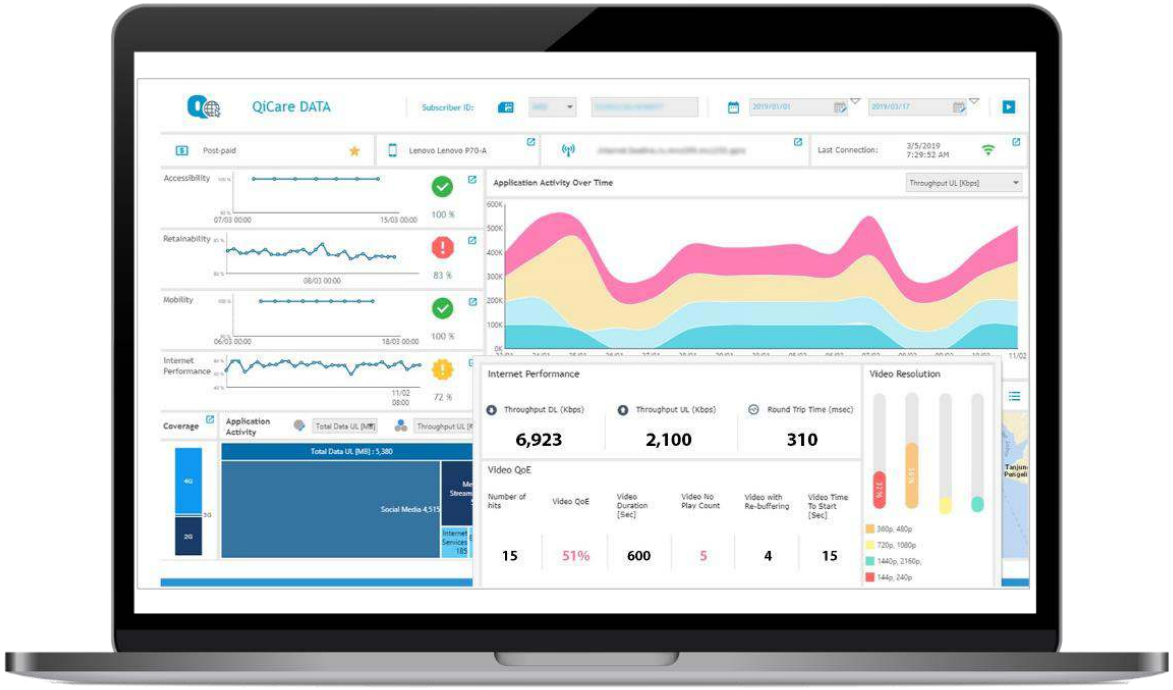


Figure 10 - Using AI/ML RADCOM provides insights into encrypted video streaming services

Operators recognize that video streaming is a potential area for growth and often offer their streaming platform to monetize and sell alongside existing video streaming services. However, to provide and monitor these services, the operator must gain critical insights into an almost entirely encrypted area.

RADCOM's solution provides operators with an understanding of the perceived Quality of Experience (QoE) for HTTPS and QUIC based video streaming such as Netflix, YouTube, Facebook, Amazon Prime, and regional specific services (like Iflix and Viu). With these insights, operators can gain visibility across encrypted networks and understand the QoE across their network, critical. Different regions are affected in different ways, and traffic usage varies. The data provided by RADCOM's solutions enable the operator to understand whether video streaming is a high enough quality that can be consumed and enjoyed by customers while not overburdening the network.

RADCOM's probe aggregates encrypted video sessions

- Identifies video stream for encrypted HLS, MPEG-DASH, YouTube QUIC, and more
- Identifies overlaying video applications such as Amazon, Facebook, Netflix, and YouTube
- Provides Video KQIs as well as TCP, UDP, and HTTPS KPIs

Provides encrypted Video KQIs

- MOS-B video quality
- Start delay
- Effective and network video payload bytes and packets
- Rebuffering indication
- Effective video throughput

RADCOM provides the following metrics for encrypted adaptive streaming:

- Minimum time to play – heuristic approach to estimate the elapsed time until the video starts to play
- Video rebuffering count and duration – AI heuristic modeling is used to estimate real occurrences of the rebuffered video segments that are detected in the network
- Video resolution duration – The video segment download duration based on four resolution levels measured in the segment level (mSec)
- RADCOM estimates the video resolution as a time series problem and considers the user's historical download data to provide results with the highest accuracy levels. The four-level video resolution durations may be used to calculate each tier's relative duration during the streaming video.
- Video Duration and byte count – duration of the video and DL byte count for the played video

Video AI machine learning process

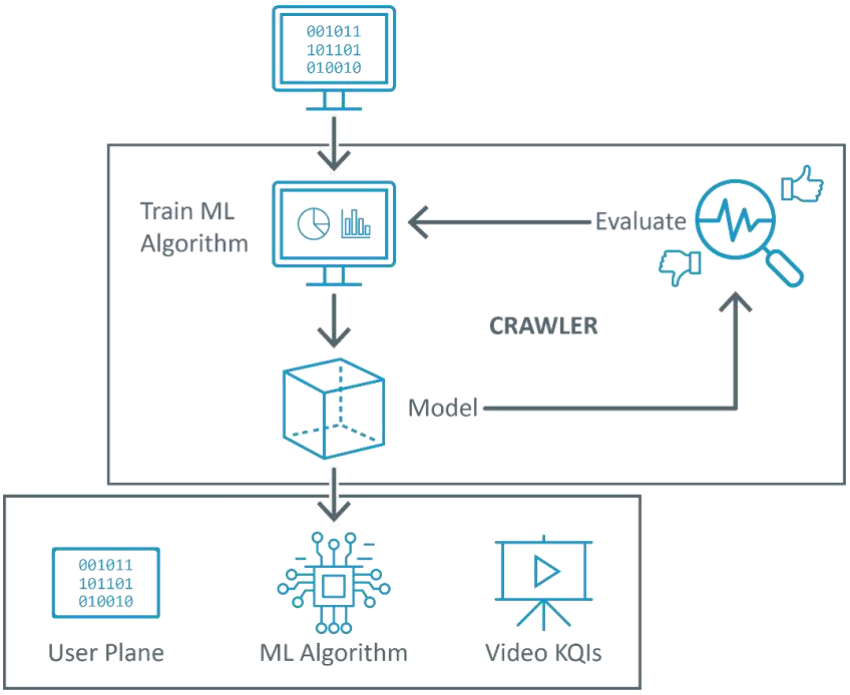


Figure 11 - RADCOM AI training process

RADCOM has developed an innovative 'video crawler' that collects network video streaming traffic, user experience, and application feedback. RADCOM's AI machine learning was trained on the data extracted from the user experience feedback and the application data used as data labels (supervised learning). RADCOM then correlates the data labels with the network traffic to calculate video KQIs from the network traffic behavior only. The training phase is performed offline under diverse network traffic and conditions from the operator's network alongside the RADCOM network lab. The model is created and validated against historical data traces (labeled data) and user test data.

For gaming

Gaming requires a premium internet connection and adds value to the operators' content offering. So, as operators move to 5G, they will offer customers a consistent and predictable Quality of Service that guarantees low latency. To ensure this service offering, operators will need to have a dedicated assurance offering that monitors these requirements, assures the network, and delivers on the demanding gaming SLAs.

RADCOM's solution offers operators the following benefits:

- Reduces E2E MTTR to identify/pinpoint areas or individuals with poor gaming latency performance
- Intuitive and useful RCA capabilities to carry out efficient and proactive E2E resolution

- Improves the accuracy of detecting high traffic sites due to exceptionally heavy users or any abnormal/massive devices, thus reducing turnaround time
- Provides insights into top gaming application trending resulting in new revenue streams
- Improves customer support for gaming latency issues
- Guarantee SLA for gaming plans

Using machine learning and heuristic modeling, RADCOM can detect patterns in the data, setting KQI's and enabling visibility for the operator where it was previously a blind spot. RADCOM can monitor and analyze the service experience segmenting via location, network element, application, device, VIP groups, and roaming partners to provide a quality experience for gamers. The end-to-end view enables the operator to pinpoint areas or individuals experiencing high latency and perform drill-downs with troubleshooting on-the-fly. Operators can identify the top gaming applications and target customers, accordingly opening up new revenue streams alongside plans for guaranteed latency levels.

Gaming KQIs provided by RADCOM's solution:

- Throughput
  - For all gaming Apps
  - DL and UL effective and peak throughput calculated for multiple concurrent TCP/UDP flows
- Protocol handshake initial and intersession latency
  - Calculated for identified gaming protocol handshake (TCP, SSL, QUIC, etc.)
  - Initial RTT based on a handshake at the start of a flow
  - Min/Max and average RTT during the TCP flow based on the elapsed time between large packets and corresponding accumulative ACKS
- Game Control RTT – Using Machine Learning
  - Games, such as Fortnite, send regular control updates between client and server
  - A high update rate may indicate poor game performance
  - Control RTT calculated as the elapsed time between subsequent client and server control update messages (calculated using machine learning)
- Quality Metrics for Gaming Sessions
  - Latency calculated for each TCP and UDP/RTCP flow
  - Aggregated gaming session latency calculated from all gaming TCP and UDP/RTCP flows
  - Periodic metrics for long gaming sessions

For tethering

Many smartphones and PC devices connect to the internet via a host device. For mobile networks, this could be a hotspot or smartphone. This would be via a Wi-Fi router for fixed-line broadband, and there may be other devices connected via a Wi-Fi or Bluetooth connection. The challenge for operators is to understand how many devices may be tethering to the host device to gauge the volume of traffic being carried through tethering.

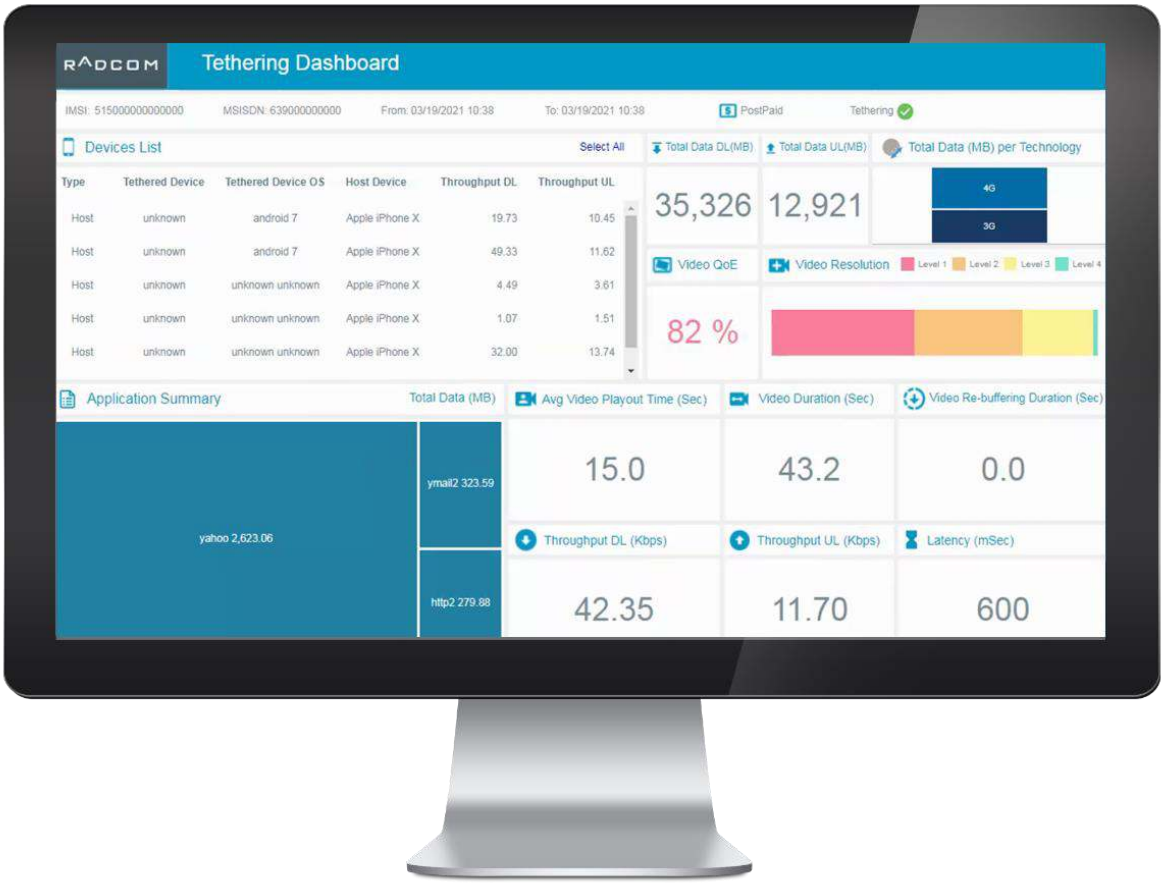


Figure 12 - Gaining insights into tethering usage on the network

RADCOM employs ML and heuristic modeling, identifying where the traffic originates from and what type of traffic it may be, whether mobile, fixed-line, or Wi-Fi/ Bluetooth. The operator can then tailor the solution to the user behind the tethering device and offer them a personalized plan to encompass their needs. Also, operators can see which Over-The-Top (OTT) applications are being used to tether, providing further insights on what demands are being placed on the network and customer usage patterns.

RADCOM's tethering solution offers insights for a single subscriber, including when they activate tethering, the number of tethered devices, total data, session duration, speed, and the Quality of Experience (QoE) for all tethered devices.

Conclusion

To ensure 5G network quality, operators will need to deploy cloud-native assurance solutions with built-in AI/ML. This will provide the most efficient way to embed artificial intelligence into the network that will help engineers manage their networks and separate the network “wheat” from the chaff automatically. This will free engineers up to spend more time on the critical task of optimizing the network performance and solving network degradations to ensure superior customer experience and operational excellence in 5G. It also lays down the foundations for an open/closed-loop approach to managing network operations.