SERVICE ASSURANCE IN THE 5G AND CLOUD ERA

Time to abandon the geocentric model. Many service assurance solutions are based on old device-centric principles. In this paper we show a scientific and proven method to achieve a proactive service-centric operations model with the customer in focus.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The 5G and cloud era are upon us and there is an industry consensus that network automation and new service assurance model are paramount to achieve the level of experience that customers expect today.

In yesterday's physical networks topologies were static, services were not provisioned dynamically, and they almost never changed. Business applications were not very demanding, and they did not require strict network KPIs. The end-to-end data plane topology and quality was well known.

Today's network services have a totally different level of complexity:

- Overlay networks based on SDN techniques make the actual topology virtual and constantly changing.
- Network virtualization and telco clouds are used to deliver distributed intelligence and value-added services in the multi-access edge network.
- Dynamic service provisioning implies a new rate of change never seen before.
- Demanding 5G mission-critical services require proactive real-time SLA monitoring.
- 5G slices assume guaranteed quality of service (QoS), which needs to be tested and monitored.
- Network functions virtualization (NFV) service chains need to be validated at the data plane to ensure that service chains provide the required KPIs. It is not enough to only monitor the health of the virtualized network functions (VNFs).

As 60% of network problems are still discovered first by end users or not reported at all, traditional service assurance approaches have proven to be failing even with the most simplistic network service architectures. How can we then deliver quality over dynamic service offerings and modern connectivity options such as SD-WAN and segment routed VPNs? How can we support 5G ultra-reliable, low latency and massive IoT communications service level objectives while networks are becoming highly dynamic and even more complex?

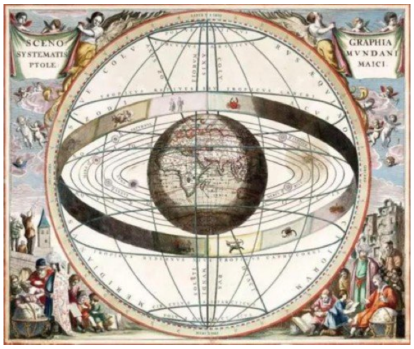
Assurance solutions have been stuck with using traditional monitoring systems which lack service quality insight and are by nature reactive. Assurance solutions must change the focal point from the infrastructure to actively measuring the actual service quality. This paper explains how to do that with two foundational principles of active assurance:

1. We need to measure at the data plane.
2. We need to use synthetic traffic to be proactive.

This means we need to transform from a model where we study the devices and infrastructure to one where we study the quality of the actual network services. To better understand this choice, this paper presents an analogy based on a geocentric to a heliocentric model. Read this white paper to learn what active assurance is and how it fits into today's overall service-centric operational landscape.

Introduction: From a Geocentric to a Heliocentric Model

Service Assurance Today



For the major part of the history of mankind, the geocentric model of the solar system was the accepted truth, where the earth was in the middle with the sun, moon, and other planets orbiting around it.[1] Today, we all know that the true configuration is the opposite—a heliocentric one with the sun in the middle.

The transition to this truth was not an easy one and required people at the time to approach the problem differently, looking at reality from a new perspective.

Today, a majority of the telecom industry considers infrastructure health as the accepted truth for service assurance. With more than *60% of network problems being first reported by customers* [1], it is about time to approach the problem differently and from the customer's perspective. The need for a change of perspective is even more evident when realizing that telecom is the worst performing industry when it comes to Net Promoter Scores (NPSs) [2]. Service providers who want to stay competitive need to shift from a network infrastructure mode of operation to a digital experience focus, moving to customer- and service-centric operations.

Gaining a New Perspective—Like Galileo

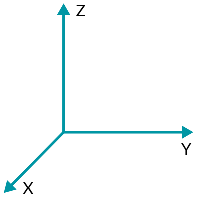
"All truths are easy to understand once they are discovered;
the point is to discover them."

– Galileo

Back in the 16th century, when Galileo was observing the planets surrounding us, there were signs in their movements that indicated that the assumption of a geocentric planetary system could not be correct. To fit Galileo's new observations into the prevailing geocentric model, more and more elaborate exceptions had to be added to the model to keep it correct. Galileo, armed with a new perspective as well as a telescope, approached the problem differently and defined a new model with the sun in the middle. This new approach gave clear explanations and predictions of the movements of all the planets. The heliocentric model was born.

As we all know, Galileo was questioned and even brought to inquisition. Today, his theory has been proven to be the immutable truth and acts as our frame of reference. [2]

The Importance of a Correct Reference Frame



Models are defined to give answers to questions. For example, models of the sun and planets should predict the movements of those bodies. Models of this kind are based on a frame of reference, "a structure of concepts, values, customs, views, etc., by means of which an individual or group perceives or evaluates data, communicates ideas, and regulates behavior." [3] Think of this as the axioms of the model. If you approach any problem using the wrong reference frame, a clear sign that you are on the wrong path is that you are constantly forced to define exceptions and apply cumbersome techniques in solving the problem. Just think of the medieval scientists trying to predict the planetary and lunar movements based on the geocentric model: the model got more and more complex the more precise predictions were needed.

[1] Ptolemaic diagram of a geocentric system, from the star atlas Harmonia Macrocosmica by the cartographer Andreas Cellarius, 1660, https://www.britannica.com/science/geocentric-model
[2] Long-Lost Letter by Galileo Shows He Tried to Trick the Inquisition, https://www.history.com/news/galileo-letter-trick-inquisition-earth-sun

A relevant question in telecom is:

- Are your customers happy with the service quality you deliver to them?
- Can you answer this question with precision? In real time? Proactively?
- Do you invest too much in research projects to get the answers?
- Do these projects really deliver results and business value?

Only if you approach the problem from the right perspective and with the right frame of reference will the model become small and beautiful, without overengineered and complex procedures and methods. Often, your solution requires only a small and focused mathematical toolbox that gives high value using a minimum of effort.

Read on as we explain active assurance using the heliocentric model, which gives the answers to fundamental questions on service quality simply and elegantly.

Introducing the Heliocentric Model for Assurance

It might be bold to state, but in some circumstances the service assurance industry has not evolved fast enough with the modernization of network technologies that have enabled an increasingly fast rate of change towards new and highly dynamic services. The industry still believes in the geocentric model, where we use the information about the infrastructure, devices, and VNFs as the frame of reference. Impact on network services and service quality are viewed as side effects. We try various complex approaches to infer the quality of the services based on what we are able to observe from the devices. This reminds us of the old geocentric model (Figure 1), as this approach is trying to solve the problem using an incomplete perspective and assumptions.

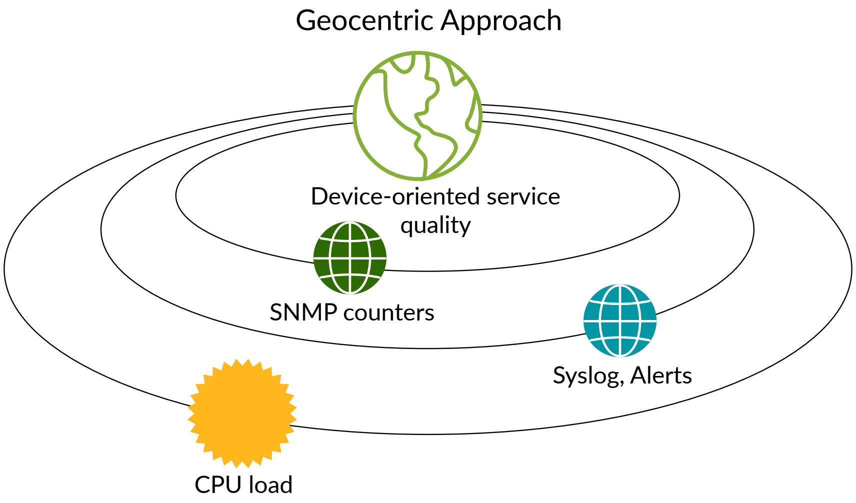


Figure 1: The geocentric approach

In the cloud and 5G era, successful orchestration platforms all have a common theme: services are first-class citizens and the underlying network with orchestration enables a service-centric approach. Orchestration possesses a complete view for service operations and manages the service life cycle. Device/VNF configuration constitutes side effects of the service where intent is programmed into the orchestration and control software layers—and not the network infrastructure itself since the network does not have end-to-end visibility of services and their business and operational intents (simplified in the figure below as "service intent").

We illustrate this with a schematic diagram (in Figure 2) that is commonly used in conferences on service orchestration.

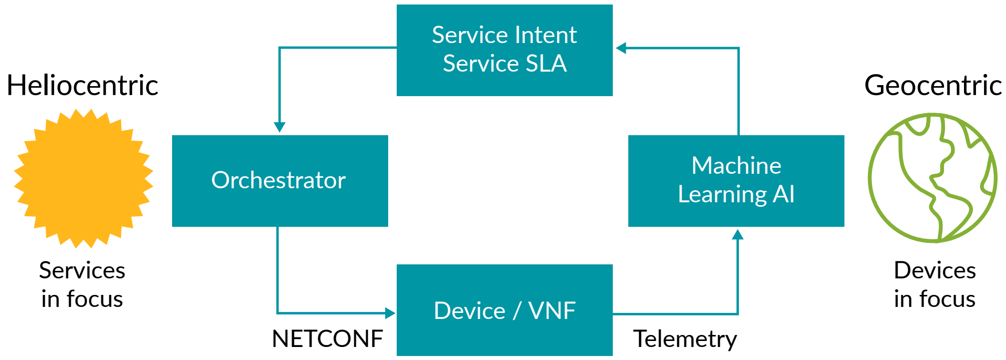


Figure 2: Today closed-loop orchestration model, mixing heliocentric and geocentric models leading to incomplete service quality insights

To the left, we see a heliocentric approach in that device configuration is derived from the service intent. But to the right, we see a geocentric model in that we try to infer service health from device telemetry. The latter is a very complex task and to this date not proven to be tractable. Even with the evolution of machine learning and big data technologies, it is extremely difficult to get real-time service insight into, for example, network key performance indicators (KPIs) like loss, latency, and jitter based on device telemetry. We need to have the same frame of reference for both provisioning and assurance.

A modern, symmetric architecture is illustrated in Figure 3.

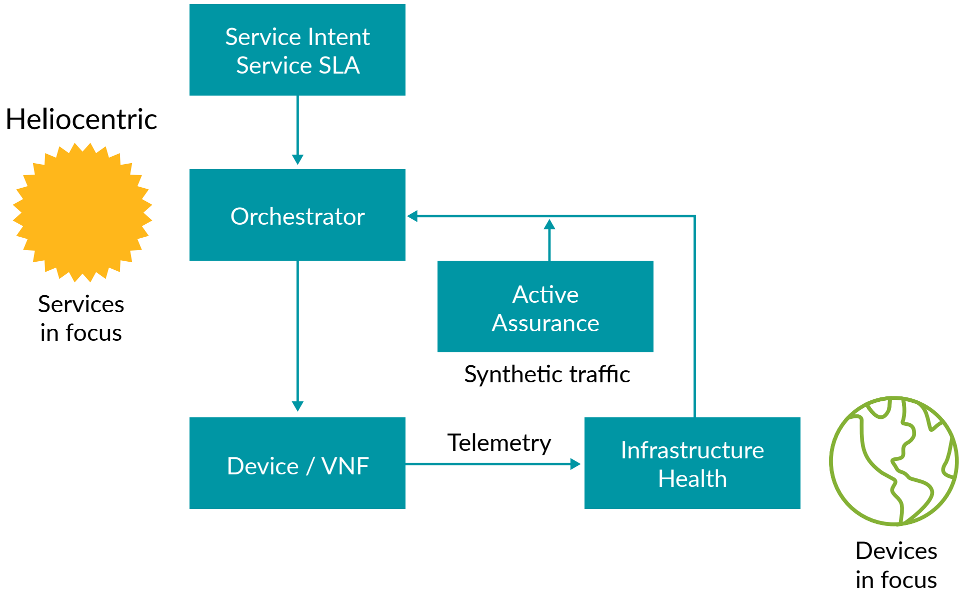


Figure 3: A symmetric heliocentric model for service orchestration

This is the heliocentric model. The orchestrator uses a service-centric frame of reference for both provisioning and assurance. The new component—active assurance—is added as the main method of measuring service quality. Active assurance is a precise technique, based on sound engineering and mathematics, for emitting synthetic traffic to measure the network KPIs with high precision and in real time.

> *“Measure what is measurable, and make measurable what is not so.”*
>
> – Galileo

As Galileo states, we need to measure to get knowledge. If we cannot measure it, we need to make it measurable. This is where synthetic traffic comes into play. With synthetic traffic, you can measure at all network layers, and you can measure from the customer's location and point of view. This cannot be achieved with device telemetry methods or passive probing. As indicated by Figure 4, the active assurance component adds real knowledge on the service KPIs like packet loss, delay, and jitter. In this model, device telemetry is primarily used for understanding infrastructure health.

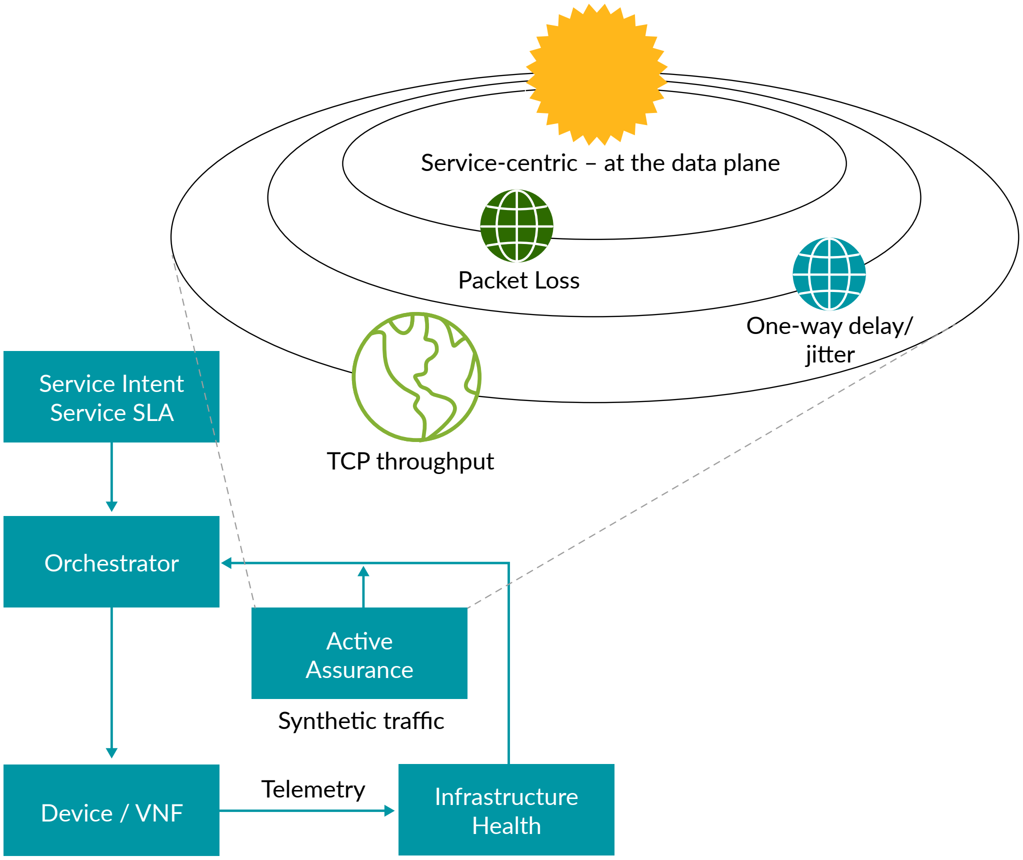


Figure 4: Heliocentric service orchestration

What About AI and ML?

So where did artificial intelligence and machine learning (AI/ML) go in the model outlined above? The service assurance industry today has high hopes for AI and ML when it comes to providing service insight. And AI and ML can certainly be useful in well-defined specific use cases. For example, if you measure KPIs like loss, latency, and jitter, you can use AI techniques to perform anomaly detection and KPI prediction. Clustering techniques can be used to group alarms. These are all examples of specific or narrow AI, the ability to act on a specific task.

However, the expectations need to be balanced as discussed in recent publications on this topic [4], [5]. AI/ML are mathematical methods that are useful in well-defined cases—given the right input data, and given an output function that can be trained. The heliocentric approach is an enabler for successful implementation of your specific AI solutions; you need high quality service-related data for your algorithms. You cannot base ML attempts to determine service quality on data lakes holding low-quality device data.

Geocentric Inertia

Inertia is always in the way of change. Now you probably object: "We have talked about network and service operations center transitions for decades. This is nothing new." But if we apply empirical methods and evidence to our existing assurance processes and tools, we find that they are still device- and VNF-focused. These are the facts:

- We have fault and performance management systems that ingest telemetry from the infrastructure.
- We try with various techniques to infer the service quality, ranging from inventory lookups and big data analytics to machine learning. All of these are useful techniques but will not give us 100% insight into real-time service quality.
- The assurance solutions are getting more and more complex, but they are still not answering the question: "What is the real-time service quality of every service I have sold?"
- Marketing promises of AI being able to transform device telemetry into service quality are not well rooted in scientific proof or empirical evidence.
- The assurance industry is stuck in incomplete geocentric frame of reference, according to which assurance is a central system that collects device data and does whatever it can with that data.

In a later section "Debunking Active Assurance Myths," we will provide you with an eye-opening summary of many misconceptions about active assurance.

"I think that in the discussion of natural problems, we ought to begin not with the Scriptures but with experiments and demonstrations."

– Galileo

The Telescope for Service Quality: Active Assurance

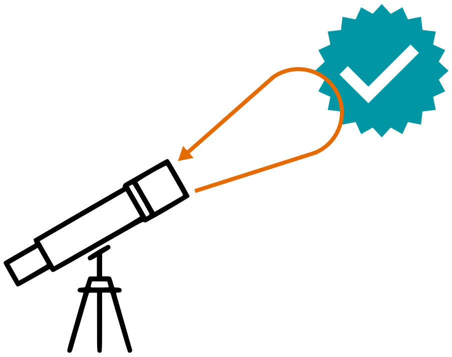


Figure 5: The service quality telescope

A critical driver behind the heliocentric model was that Galileo used the telescope. By observing the phases of Venus, he was able to conclude that data was inconsistent with the old geocentric model but readily explained by the heliocentric one. We have been trying for a long time to understand service health by sensing device data. This has shown to be complex and incomplete.

The problem for the service assurance industry is that we do not think there is a direct way to measure what we are looking for: true service quality. The traditional assurance approach is to use a model that assumes that network services are something abstract, that they do not really exist, and that they cannot be directly monitored. This axiom has proven to be incorrect with the advent of "active assurance" (discussed in more detail further on within this paper).

The whole purpose of network services is to transfer packets between clients and servers across a network. This network can be more or less complex. Figure 6 illustrates a typical network today, spanning cloud, SD-WAN, mobile, campus, and Wi-Fi segments.

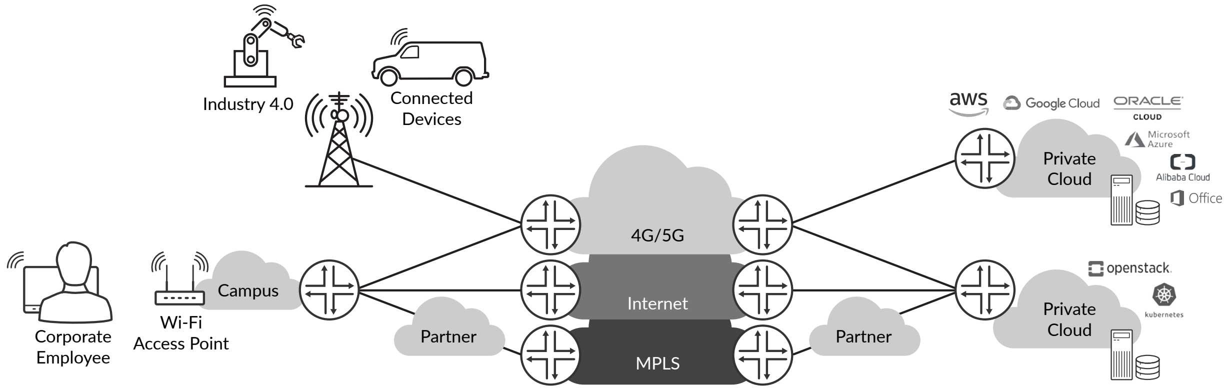


Figure 6: A schematic network diagram of today's typical network

The characteristics of this packet transport are what matters. Customers directly perceive these characteristics. As a service provider, this is what you sell, what you charge for, and what your business depends upon. This applies to business services, mobile backhaul services, cloud services, and residential services. Customers do not perceive statistics available at the management plane from a number of devices.

In a modern service assurance approach, there is a telescope that uses a frame of reference based on synthetic traffic to directly measure the service quality on the data plane. By measuring KPIs for synthetic packets that travel on the same path as customer traffic, end to end, we get direct insight into the service quality. This is what matters for the customer. And since the traffic is synthetic, we know beforehand. We do not have to wait until the customer experiences an issue, or outage [6].

Let us first explain "active assurance," (see Figure 7 below) since this concept might be unclear. It is based on the following:

- Small virtual test agents that emit synthetic traffic and measure against defined test outcomes and SLAs, acting as virtual customers of the services
- Test agents using standardized reflection technology, for example Two-Way Active Measurement Protocol (TWAMP) [7] and Y.1731 [8]) built into most network devices
- Central control, data collection, and analysis of the measurements

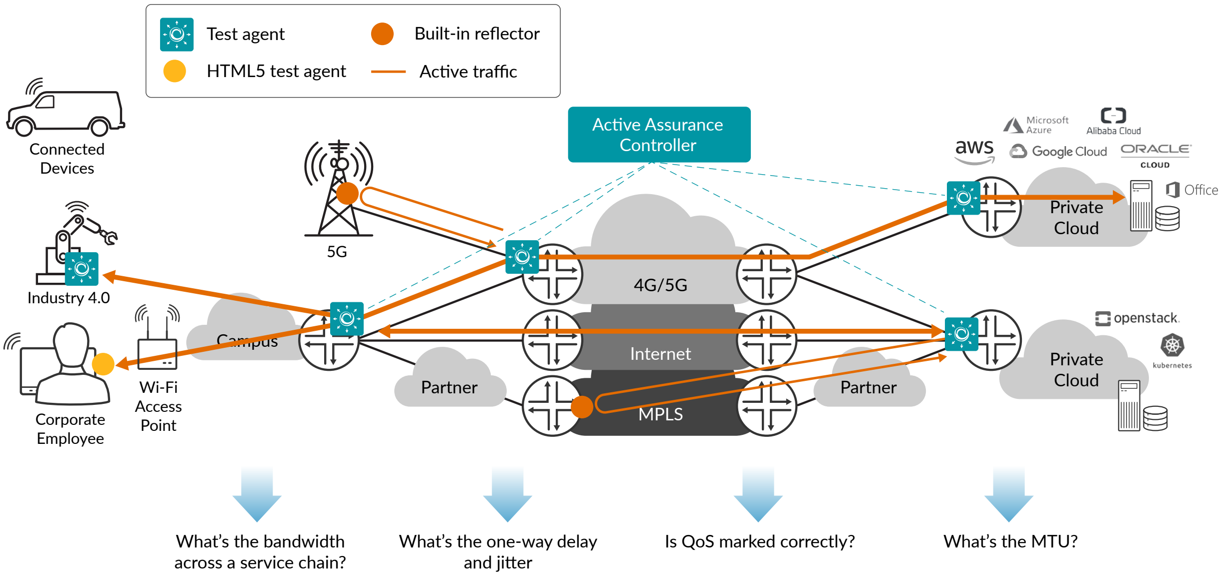


Figure 7: Active assurance across a network

The modern, heliocentric assurance model puts the customers in focus by measuring the service quality, the same way as customers do. Once you start thinking in this way, it becomes obvious. You are selling services to customers. You will be judged on how customers perceive the services, not on how individual devices behave. How can you measure this? The only true way is to test and monitor as a customer.

After services have been deployed but before customers start using the service, test agents perform time-limited intrusive tests to validate load scenarios. This results in an activation test report that confirms what is sold really works. Later, when the service is operational, the same test agents emit small numbers of packages (a few packets per second) into the network in order to continuously validate technical SLAs and KPIs like Mean Opinion Score (MOS), loss, latency, and jitter. This is virtually nonintrusive monitoring.

It is, of course, impractical to deploy agents all across the network. It is a bit unknown to the market that built-in reflection technology on L2 (Y.1731) and L3 (TWAMP) is standardized and widely supported by vendors. This means that you can have a few test agents installed centrally that will perform active measurements against endpoints using built-in reflection techniques.

An active assurance controller drives test cases and ongoing monitoring. The controller manages test and SLA definitions and instructs test agents to send synthetic traffic to run tests and validate SLAs. Another fundamental task of the controller is to manage the life cycle of the test agents.

The test controller supports automation APIs so that it can act as a component within the overall orchestrator.

See more on active testing in an analyst report by Analysys Mason: "Active testing techniques can help communications service providers to improve the customer experience" [9].

Benefits of Active Assurance

There are several important business benefits of using active assurance:

- You can test services before they are used as part of delivery/orchestration.
- SLA monitoring of data plane KPIs provides your service operations center with real-time insight into service quality with high resolution of loss, jitter, delay, response times, and multiple measurements every second, down to microsecond resolution.
- It is easy to deploy. No integration towards the management plane is required, no MIBs, no telemetry schemas, no vendor adaptors.
- It can be easily orchestrated/automated. This is fundamental to cover use cases such as service chains and slicing.
- Mature technology and products are available
  (e.g. Juniper® Paragon Active Assurance has been proven in 200+ customer deployments).
- Your customers will not be the first to report performance degradation issues.
- You can pinpoint third-party providers when services span networks that are not fully owned and managed by you. Active assurance will show within seconds if your partner is not delivering the promised quality.

Debunking Active Assurance Myths

Looking at the benefits above, you might ask yourself: If this technology is available and the claims made are true, "Why isn't active assurance the primary tool in our service operations center?" The answer is much investment has already been made with the perception of expensive cost for this value. However, active assurance is not like traditional probe based solutions because it is more cost-effective and achieves significantly better results.

Probes and probing are overloaded terms. A common barrier to adoption is that heavy investments have already been made in passive probing solutions, and management is often unwilling to invest in additional probing solutions. However, as explained in earlier sections, active test agents (active probes) solve different problems than passive probes, by using synthetic traffic to measure quality the same way as end users. In contrast to classical passive solutions, active assurance solutions scale from small to large networks in both pricing and automation.

In addition, there are many myths in the industry about active assurance, and those resistant to augmenting and investing in more modern solutions sometime use these as excuses not to implement it. The table below lists some of the most common misconceptions:

Myths	Reality
Active methods require test agents everywhere.	Test agents can be deployed with a low total cost of ownership on a few locations centrally and combined with reflection methods like TWAMP that are built in many devices.
Test agents are expensive.	Today, test agents are provided as a cost-effective software.
It is very complex to deploy active test agents (active probes).	Zero-touch methods allow for large-scale, fully automated deployments, anywhere from universal customer premises equipment (uCPE) to the cloud without any humans in the loop. It is simple to automatically deploy thousands of test agents in just minutes.
Active test agents require lots of compute resources.	Test agents are small VMs or containers, requiring only a fraction of a virtual CPU (vCPU) and less than 100 MB of RAM. This makes them suitable for deployment even on network devices, which typically have limited capacity for application hosting.
The only way to get service quality insight is through passive probing of existing traffic.	Passive probing is exposed to a serious blind spot when no traffic is seen. Is the service broken or is the customer not using the service at the moment? Using active traffic, you know what you generate and what to expect. Central passive probing lacks true end-to-end service quality visibility as traffic is only captured at central locations, and not at the edges. Also, when overall traffic grows exponentially, so does the cost for passive solutions to keep up with processing the increasing amount of captured traffic. Active methods ride on top of other traffic, with no similar cost driver. Finally, passive probing is only possible after launching the service, without the chance to test and validate the service before going into production.
Service quality can be accurately inferred by collecting data and statistics from my network devices.	If your goal is to monitor real service quality, you need to actively consume services on the data plane, the same way as end users, instead of looking at device data collected from the management plane (SNMP, log files, and telemetry), which is focused on infrastructure health.
Active monitoring with synthetic traffic will load my network and consume all my bandwidth.	On a 1 Gbps link, when using 64 byte packets sent every 100 ms (10 pps), the synthetic test traffic would require only 0.0005% of the total capacity. This means that active monitoring traffic is essentially negligible compared to other traffic, even when multiple monitoring sessions are used. In addition, unlikely passive probing, active assurance can also be scheduled to perform testing during low traffic periods time such as weekends or typical sleeping hours when then network utilization is lower.
Passive is better as it looks at real user traffic.	With passive techniques, you will always be a step behind your customer. You can confirm that the customer is dissatisfied, but you cannot act beforehand.
Fault management and problem management systems are good enough to build a proactive solution.	Most of the issues in a network are not faults as such, but are rather due to non-optimal configurations. And by definition, there is no device alarm for a non-optimal QoS configuration or non-optimal routing policy/traffic engineering metric. Even misconfigured firewall rules also lead to silent performance issues that are difficult to pinpoint as well. With active assurance, these types of issues become easy to detect.
Public cloud providers will soon offer passive data sources to our legacy tools.	Some offer this already today, but pricing is volume-based pricing and the approach has very limited scalability. Typically, it is intended for security use cases rather than for service assurance. In the age of digital security and cloud content, active assurance provides a solution that can work to test encrypted services when passive approaches do not.

Transforming Network Operations to Become Service-Centric

In this section, we will look at how a modern heliocentric model can transform your network operations center (NOC) to become more service-oriented. First of all, we want to make clear that we are not saying that active assurance systems can replace fault management (FM) systems, performance management (PM) systems, or passive probes. That is not the case: you still need these to understand the infrastructure health and general utilization in your network. But none of these will give you real service quality insight that will prevent the 60% of service-impacting issues and so far have only been discovered by costly customer complaints. There is a better approach by augmenting your existing investments with active assurance to proactively detect these issues before customers are impacted.

Current NOC Model

An oversimplified picture of assurance solutions in a traditional NOC is shown below:

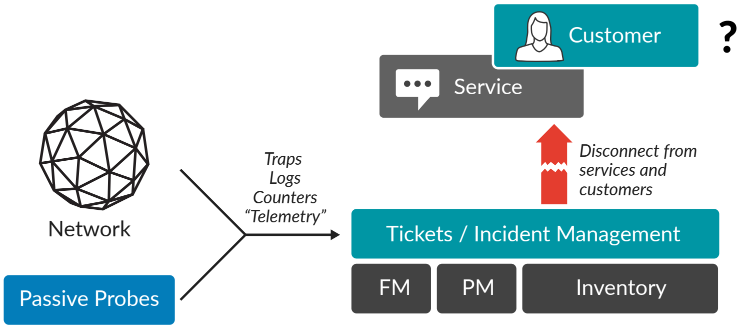


Figure 8: Traditional NOC

This model is the old geocentric one. Data collection is focused on the infrastructure. Some of the symptoms of having the wrong frame of reference are listed below. Use this list to assess your current operations center:

- Customers report many of the service-related issues before the NOC detects them.
- Operators work with alarms and tickets associated with devices, and services are abstract entities.
- To some degree, the services are represented in the inventory system, and various tools and manual processes try to map issues to the services.
- It is hard to set priorities: Which alarms are really critical from a customer perspective?
- Issues in the network depend on heroes in the NOC—the people who have the service structure in their heads and a deep knowledge of how to analyze and troubleshoot.
- There is a gap between customer care and the NOC. Customer care works with service-related issues reported by customers, while the NOC works with issues reported by the devices.
- Finally, a common issue today is that services are not properly tested at delivery. In many cases, customers detect and report quality issues in a newly delivered service, and the NOC needs to work with an incomplete delivery.

Realizing the Service Operations Center with a Modern Heliocentric Model

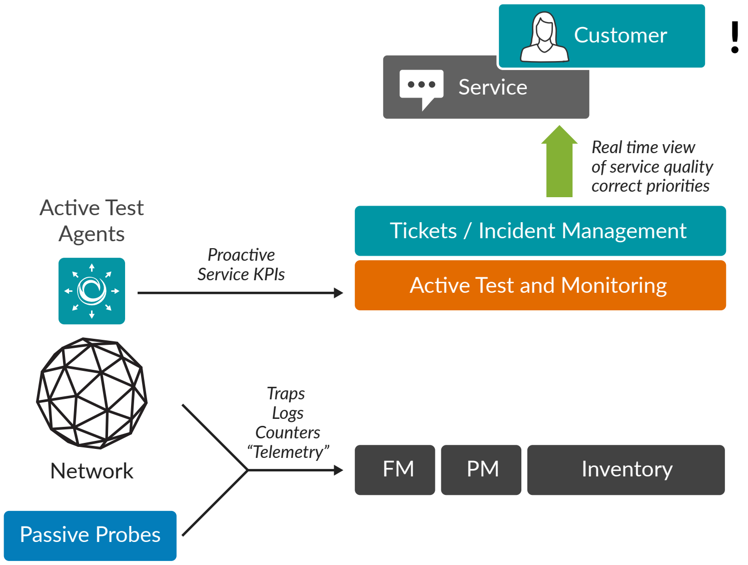


Figure 9: Heliocentric service operations center

To build a true service operations center, you need to put the services in the center, just like the sun in the heliocentric model. Then, your devices and infrastructure are secondary, just like Earth and the other planets. This means that the main screen in your service operations center should display violated service SLAs as monitored by active assurance. It should represent the services that you have sold, and they should be monitored end to end, as the customer uses them. By using this approach, you ensure that you are working on the right priorities and can act before customers do. Your classical NOC is pushed to a debugging layer where their network troubleshooting expertise will be best utilized.

Another very important aspect is that the services are all tested as part of the delivery process. The NOC will not need to debug incomplete deliveries.

Conclusion

With the 5G and cloud era upon us and automated assurance being paramount to achieving the expected level of customer service experience, a shift to a more effective model is needed to meet the service level objectives of new highly dynamic services, such as for ultra-reliable, low latency and massive IoT communications. It will not be sufficient to react to problems automatically, as they arise. Your operation team also need to become service-centric and proactive.

Think about what is most important to you:

- Is it to guarantee the service quality of what you sell to your customers, focus on the digital experience, transform to a new modern assurance model that uses ***active assurance*** to monitor services in real time?

[or]

- Is it to monitor the equipment and software you bought from vendors, sticking to the old assurance model based on collecting data from the devices and VNFs?

Changing your frame of reference is painful, but an evolution to a new model is needed to be able to meet customer service experience requirements in the 5G and cloud era.

Operational support systems (OSS) and assurance are such an established domain, including misconceptions and the effort intensive search for a nonscientific magic wand. We argue that it is more worthwhile to apply scientific knowledge to how you spend your budget. Are you willing to experiment with complex IT projects to try to infer service health, or would you prefer to apply a technique that is straightforward and can provide empirical evidence in an hour? Augmenting existing investments with active assurance is the pragmatic way forward that provides the most reliable and cost-effective at a low total cost of ownership.

Albert Einstein is widely credited with saying, "The definition of insanity is doing the same thing over and over again, but expecting different results."

If you want different results, you need to try different approaches.

Go active with Juniper Paragon Active Assurance to become a true digital service provider.

Find more at https://www.juniper.net/us/en/products-services/network-automation/paragon-active-assurance/.

References

1. Juniper Networks, Understanding Network Brownouts, https://www.juniper.net/assets/us/en/local/pdf/whitepapers/2000759-en.pdf
2. CustomerGauge, The NPS & CX benchmarks report, https://customergauge.com/benchmarks-report
3. Dictionary.com, https://www.dictionary.com/browse/frame-of-reference
4. Forbes, Struggling with Fake AI? https://www.forbes.com/sites/cognitiveworld/2020/02/28/struggling-with-fake-ai-heres-how-to-become-a-real-ai-company/#3303ae7ddcc2
5. Robotics business review, Almost 80% of AI and ML Projects Have Stalled, https://www.roboticsbusinessreview.com/ai/almost-80-of-ai-and-ml-projects-have-stalled-survey-says/
6. Juniper Networks, Jonas Krogell, "Were You Affected by Google Cloud's outage? An Analysis, and How Ongoing Monitoring Could Have Helped'
7. RFC 5357, A Two-Way Active Measurement Protocol (TWAMP), Jozef Babiarz and Roman M. Krzanowski and Kaynam Hedayat and Kiho Yum and Al Morton, https://www.rfc-editor.org/rfc/rfc5357.txt
8. Y.1731, Rec, ITU-T Y. 1731, OAM functions and mechanisms for Ethernet-based networks
9. Analysys Mason, Anil Rao, Active testing techniques can help communications service providers to improve the customer experience, https://www.analysysmason.com/Research/Content/Comments/active-testing-techniques-rma01/

About Juniper Networks

Juniper Networks brings simplicity to networking with products, solutions, and services that connect the world. Through engineering innovation, we remove the constraints and complexities of networking in the cloud era to solve the toughest challenges our customers and partners face daily. At Juniper Networks, we believe that the network is a resource for sharing knowledge and human advancement that changes the world. We are committed to imagining groundbreaking ways to deliver automated, scalable, and secure networks to move at the speed of business.

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

APAC and EMEA Headquarters
Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk
Amsterdam, The Netherlands
Phone: +31.0.207.125.700

Engineering
Simplicity