AI intent-based operations: The key to building autonomous networks

How implementing a service intent manager can benefit service assurance

ERICSSON

Bringing autonomy to your network through intent-based operations

Contents

Implementing a service intent manager for artificial intelligence (AI) intent-based operations is an important foundational step in the journey toward autonomous network operations.

As the telecommunications industry becomes ever more complex, service providers need to be able to manage the challenges involved in delivering tailored 5G differentiated connectivity, while continuing to meet the strict service level agreements (SLAs) they already have in place.

With the advent of new technologies such as 5G network slicing, there is a need for service providers to deploy adaptive automation. One solution is AI intent-based operations, which enables machines to manage conflict within multi-layer 5G technology, dynamically handling slices and partitioning traffic to meet business requirements. This improves agility and enhances efficiency by managing customers and enterprises' intents efficiently – as well as by reducing the maintenance period for upgrades.

Implementing AI intent-based operations enhances the overall level of autonomy and drives the evolution from traditional data-driven operations toward more autonomous and intent-driven operations.

The early adoption of AI intent-based operations in the service assurance area can help service providers get a head start in diversifying services and growing their revenues. However, building confidence in autonomous network operations will be a gradual process. Service providers must be confident that not only could a service intent manager be utilized for intent handling in the service assurance domain with minimal disruption, but that it could also complement their existing operations with automated execution. Service providers can achieve quick wins with the existing automated execution and policy-driven operations, while gradually implementing adaptive automation in the journey toward autonomy.

> "Our strategic partnership with Ericsson has contributed greatly to us becoming a world leader in 5G performance. Together, we implemented the world's first multi-operator core network (MOCN), which allows Malaysia's six mobile operators to access 5G RAN with their own core network without compromising any end-user experience, or network performance."
>
> **Ken Tan**
> Chief Technology Officer,
> Digital Nasional Berhad

What do we mean by "AI intent-based operations"?

AI intent-based operations are centered around providing a statement of intent of either what a service provider wants their network to do, or of what outcome is required from the network.

The demands placed on networks are rising all the time, owing to a combination of service provider and end-user requirements, increasing complexity, and stringent SLAs. Networks need to be managed faster and more efficiently, in a way that ensures performance levels can cope with the expanding difficulties of growing networks, as well as in a way that minimizes the need for manual intervention. To achieve this, the efficacy of the network management needs to scale with the growth of the network itself.

The objective of AI intent-based operations is to efficiently manage the complexities and variations of services, ultimately aiming for something resembling hands-free, autonomous service operation.

AI intent-based operations are predicated on providing a statement of intent – essentially how a service provider "wants" its network to behave, or the outcome that is required.The AI intent-based system then interprets these requirements and manages the network in the optimal way for achieving the stated goal or outcome. Once the feedback loop is complete, and the desired outcome achieved, the solution – or process – is "learned" and implemented across the network. If, however, the intent is not met, then a cognitive loop under the intent-handling process enables dynamic solution generation, and evaluation of a suitable solution, leveraging machine reasoning and utility-based evaluation methodology.

Leveraging AI, machine learning (ML) and automation technology in this way enables the network to be both more agile and flexible, able to react to changing demands and conditions in real time. This offers a more responsive and adaptive infrastructure through the dynamic allocation of network resources, ensuring a smoother user experience. Additionally, in this way the statement of intent grants the service provider full control over the network.

How AI intent-based operations reduce complexity and conflicting intents to enable delivery of SLAs:

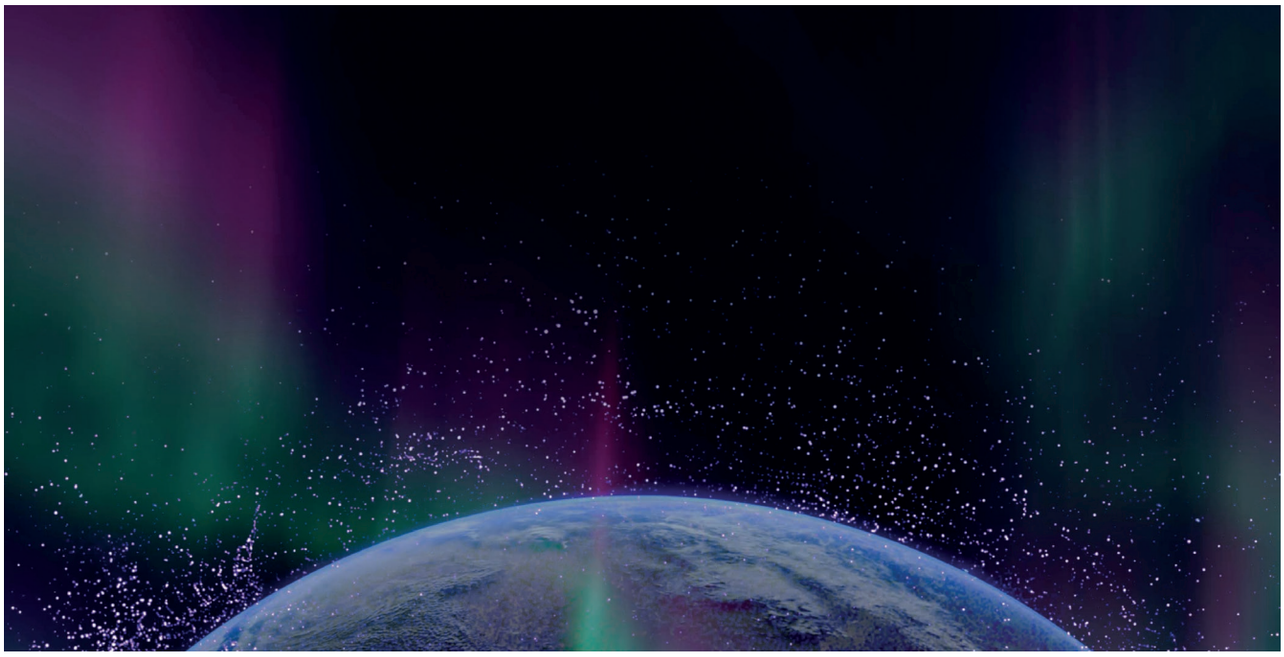
- Simplifying network management with better efficiency in network resource allocation.
- Unlocking the monetization of new services by consistently delivering differentiated connectivity.
- Allowing the network to respond, adapt and scale in a fully autonomous way.
- Offering a full and transparent end-to-end process where all decisions that the network is making on behalf of service providers are tracked.

The term "AI" covers a broad spectrum of techniques that we use to execute processes:

- **Machine learning:** Analytical techniques for data and media that allow the creation of insights, recommendations and predictions – such as performance degradation.
- **Machine reasoning:** Techniques for making decisions without preset rules, or in changing environments, such as compromise or conflict decisions, where there are competing resource demands.
- **Generative AI:** A technique that uses vast amounts of reference data to create content based upon simple language questions – such as code generation, intelligent root cause analysis or chatbots.
- **Automation:** Techniques to automate tasks and end-to-end processes to increase speed and accuracy and reduce execution costs.

Designing networks for the future

Policy-driven systems have inherent limitations in automating runtime execution.



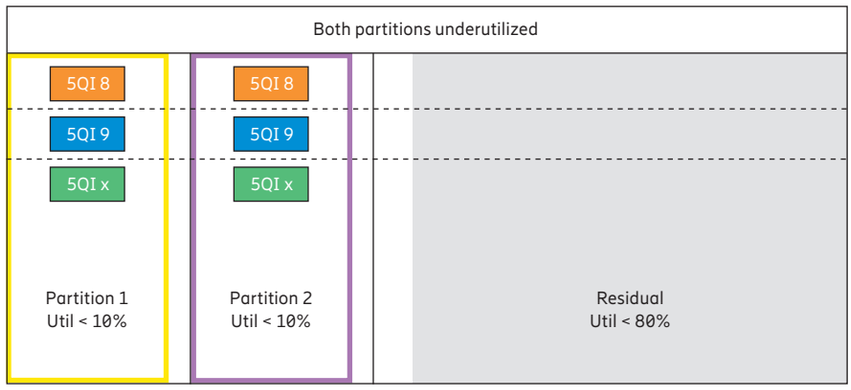
As next-generation network capabilities such as 5G network slicing and service offerings roll out at an accelerated rate and a greater scale, increasing in complexity all the while, existing automated systems will struggle to adapt to unforeseen changes and scenarios that were not explicitly considered during the design phase. This situation will lead to a need for increased manual intervention, which is ultimately unsustainable and limits the overall value of the network.

The aim of autonomous networks is to reduce human involvement in tasks such as aligning the system with business utility and strategies, as well as meeting the dynamic demands of next-generation network technologies such as 5G. In an autonomous system, intelligent decision-making is taken over by the system itself without manual intervention. The idea is to introduce AI intent-based operations in the service assurance space to provide solutions and answers to pressing – and potential – problems:

- **Problem:** Breaking away from the "complexity equals more cost" curve will require a different level of automation than we see today.
  Solution: Simplifying networks and managing conflicting interests to meet desired business outcomes through end-to-end transparent processes.
- **Problem:** There is a drive to monetize 5G due to the dramatic increase in service demand variability.
  Solution: Unlocking the monetization of new services though consistent delivery of differentiated connectivity services and SLAs.
- **Problem:** Operations need to deliver service outcomes, in addition to technical performance.
  Solution: Allowing the network to respond, adapt and scale in a fully autonomous manner.
- **Problem:** The speed and agility of processes supporting new, diverse services needs to be great – success will not come from trying to scale human-centered operations.
  Solution: Giving control back to service providers by guaranteeing a full, transparent, end-to-end process, where all decisions the network is making on behalf of the service providers are tracked.

All the current service assurance requirement input is initially onboarded through a human-to-machine interface, using everyday language which could translate service provider intent into a formal "ontology" – meaning the computing vocabulary, as specified by TM Forum. The onboarding process ensures feasibility of the intent and that there are no conflicts. Finally, AI intent-based operations in service assurance implement a service intent manager and knowledge base for the overall intent handling with machine reasoning.

Figure 1: Scenario A – depicting early network deployment phase where partitions are underutilized



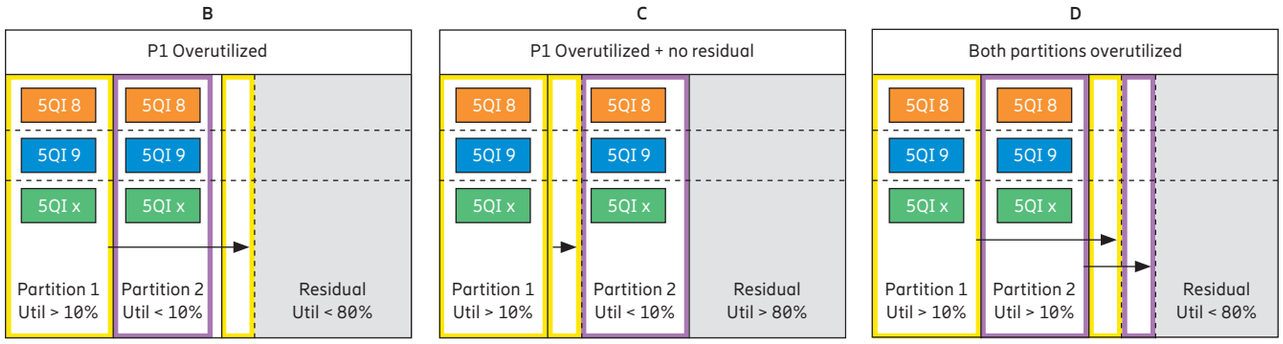
Overcoming the human limitations of automation

When more enterprise customers are added to networks, there is an increased risk of SLA violation due to radio resource contention. While the potential for automation to improve and increase value in networks is huge, in reality progress is hindered by human involvement. Partial automation levels, even at levels of 90 percent or over, will be unable to compare to the results achieved by complete autonomous operations that require no ongoing manual intervention throughout the operational loop.

However, another challenge lies in ensuring that machine reasoning can replicate human capabilities in managing conflicts and compromises.

As an example, in 5G network slicing, the radio resource partition is significant in ensuring minimal guaranteed share during radio resource contention scenarios. Partition share is minimum during resource contention only, there is no capping value to what you can use otherwise. In the early network deployment phase, all partitions are underutilized, as shown in Figure 1.

Figure 2: Scenarios during later network deployment phase where partitions are overutilized

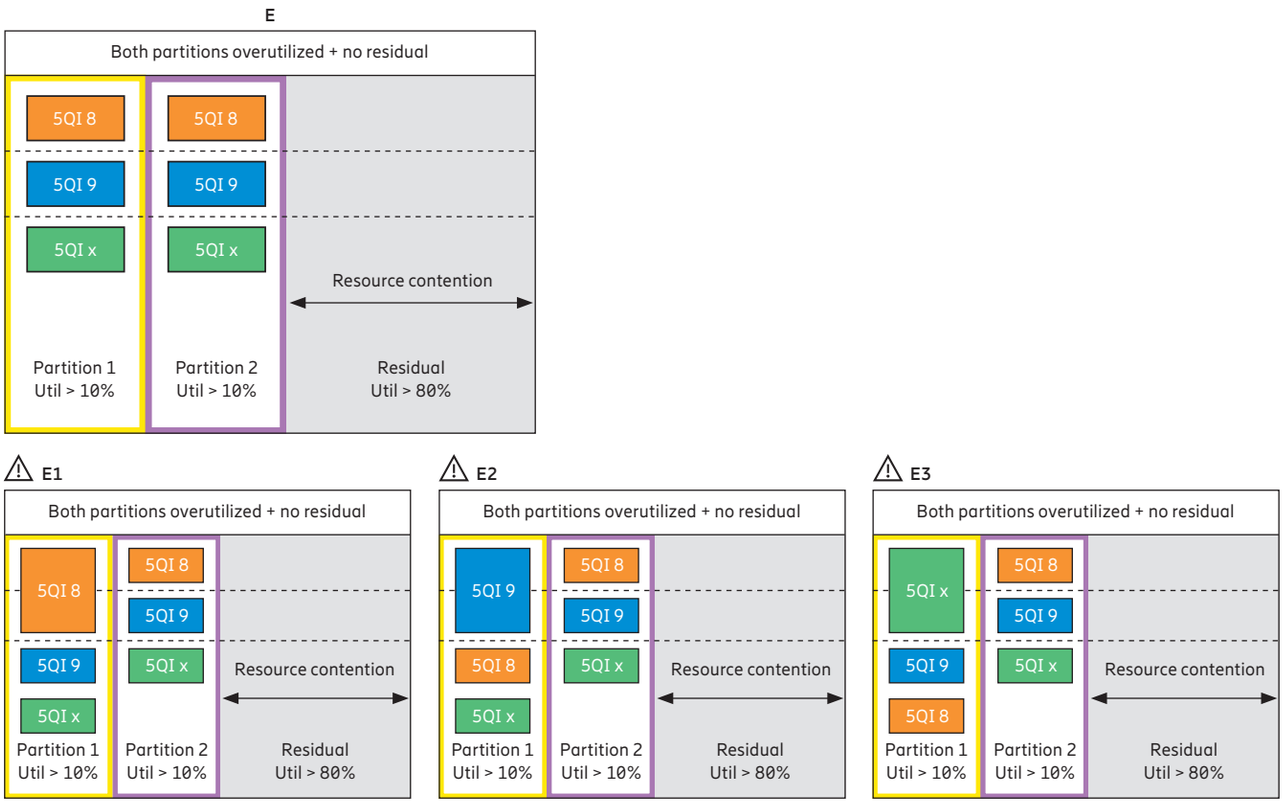


However, as the number of customers or services associated with the partition increases, there is a risk of creating a scenario where overutilization of partition will consume any spare radio resources that are available. Examples of this are shown in Figure 2, scenarios B, C and D. Therefore, the risk of continuous partition overuse associated with certain partitions exceeding the assigned partition share leads to a specific New Radio (NR) utilization challenge. This pattern of partition overuse can lead to inefficiencies in terms of radio resource usage.

Additionally, with more slices and subscribed users, there is always the risk of SLA violation due to radio resource contention. The overutilized partition will not maintain the SLA due to resource contention, as shown in Figure 3, scenario E. For example, partition may be dominated by different priority services like premium users (refer E1), normal user traffic (refer E2) or NSA traffic (refer E3).

Similarly, other dynamic network scenarios could arise from external factors including changing user behavior, factors controlled directly by the service provider, such as introducing new services, evolving energy-efficiency constraints, or customizations based on enterprise requirements. The current automated systems require human intelligence for support in order to function in such dynamic scenarios. This approach of dealing with changing requirements using manual operations support leads to inefficiencies and increased lead time. Relying on manual decisions will not be sufficient to meet the increasing dynamic network requirements for next generation networks, such as for 5G for enterprise business needs.

Figure 3: Various scenarios in E, increasing the risk of SLA violation during resource contention



The journey toward autonomy

The aim of autonomous networks is to reduce manual intervention and meet the dynamic demands of next-generation network technologies.

In an autonomous system, intelligent decision making is taken over by the system itself, without manual intervention. The system can generate new solutions and evaluate actions based on a range of concerns, even in new and unfamiliar situations. This requires system access to information such as relevant goals, requirements, targets and constraints, as well as using knowledge about utility to dynamically adapt its behavior. The automated reasoning processes enable this dynamic adaptation within a set of human-defined parameters.

AI intent-based operations translate business intents into network actions. But how does this actually work?

- Intents for the service/slice are defined and then automation, machine learning and machine reasoning ensure that the service meets a guaranteed SLA that the consumer or enterprise has purchased.
- Machine learning agents monitor the intents' key performance indicators (KPIs) and predict when an intent will be breached.
- Machine reasoning reviews the possible remedy scenarios and makes conflict and compromise decisions on resources required to meet the intents and makes the action recommendation.
- Automation is used to execute the remedial actions to maintain the guaranteed SLA.
- Humans only monitor exceptions, the process flow is fully autonomous.

Intent is the base for autonomous network operation and correlates with an increase in the level of autonomy. Mature AI intent-based operations serve as a solid foundation for building the autonomous network domains of the future. The concept is to introduce AI intent-based operations in the service assurance space so that:

- All current service assurance requirement input is initially onboarded through a human-to-machine interface, using everyday language which could translate service provider intent into a formal ontology (as specified by TM Forum). The onboarding process ensures the feasibility of the intent and removes conflicts.
- Service assurance requirements are handled without predefined rules, but instead are handled with dynamical logic and workflows with reusable components. This is much more cost efficient and has the inbuilt agility with which to meet changing business needs. The service intent manager implements the cognitive core, which in turn continuously receives observations about the environment (for example, from Ericsson Managed Services service assurances tools). The cognitive core then formulates a hypothesis based on those observations ("beliefs") and dynamically engages the required agents to determine the best actions to improve the intents ("thinking").
- All service assurance closed-loop actuation outputs are planned and managed holistically, considering all service assurance requirements, which ensures that business priorities are followed and conflicts are resolved.

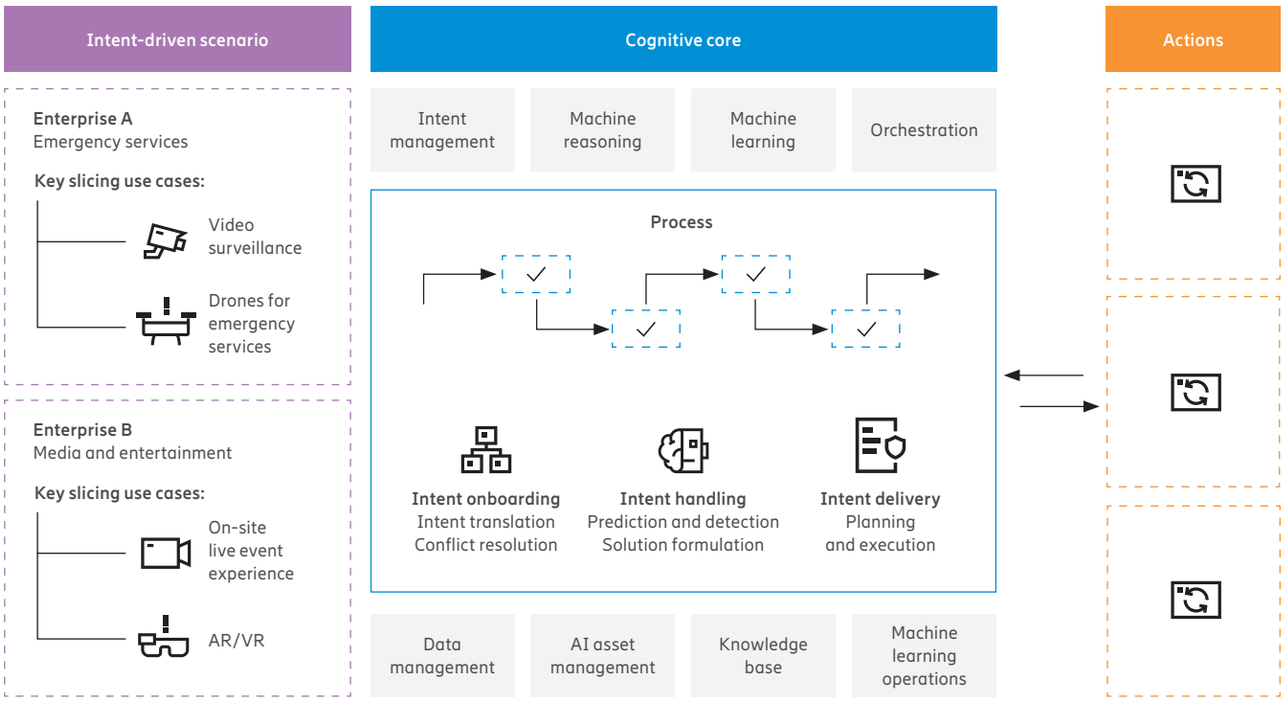
AI intent-based operations in service assurance implements a service intent manager and knowledge base for the overall intent handling with machine reasoning.

Ericsson's award-winning IBO solution

Managed Services Mastery

The best operations solution incorporating AI functionality

Figure 4: Introduction of a service intent manager and knowledge base for AI intent-based operations in the service assurance space



The knowledge base is continually updated from existing managed network services tools, like network performance management KPIs and predictive analytics from tools such as Insite. The cognitive core uses data stored in the knowledge base to:

- reason about the actions to take
- check the intents are met
- support evaluation of solutions/options with predefined strategies
- support optimized decision-making by maximizing global utility

The intent-handling loop continually runs a service assurance loop, which confirms if all the onboarded intents are met. "Intent" does not imply the start of a process; it's an imperative prescription of what action to take. It also doesn't include the prescription of specific actions, or processes to follow. Instead, intent focuses on conveying knowledge about requirements rather than dictating the actions to be taken, providing the autonomous system with the flexibility to both apply known strategies and to explore new solutions.

Building the trust in operations

The full, transparent process:

1. **The steps in the cognitive loop**
   Details every step of the cognitive flow which led to the recommendation or actuation.
2. **Adaptive automation agents**
   Explains which automation agents were engaged in recommending a specific solution.
3. **Cognitive core and knowledge base**
   Explains the core components and knowledge base contribution toward a specific recommendation.

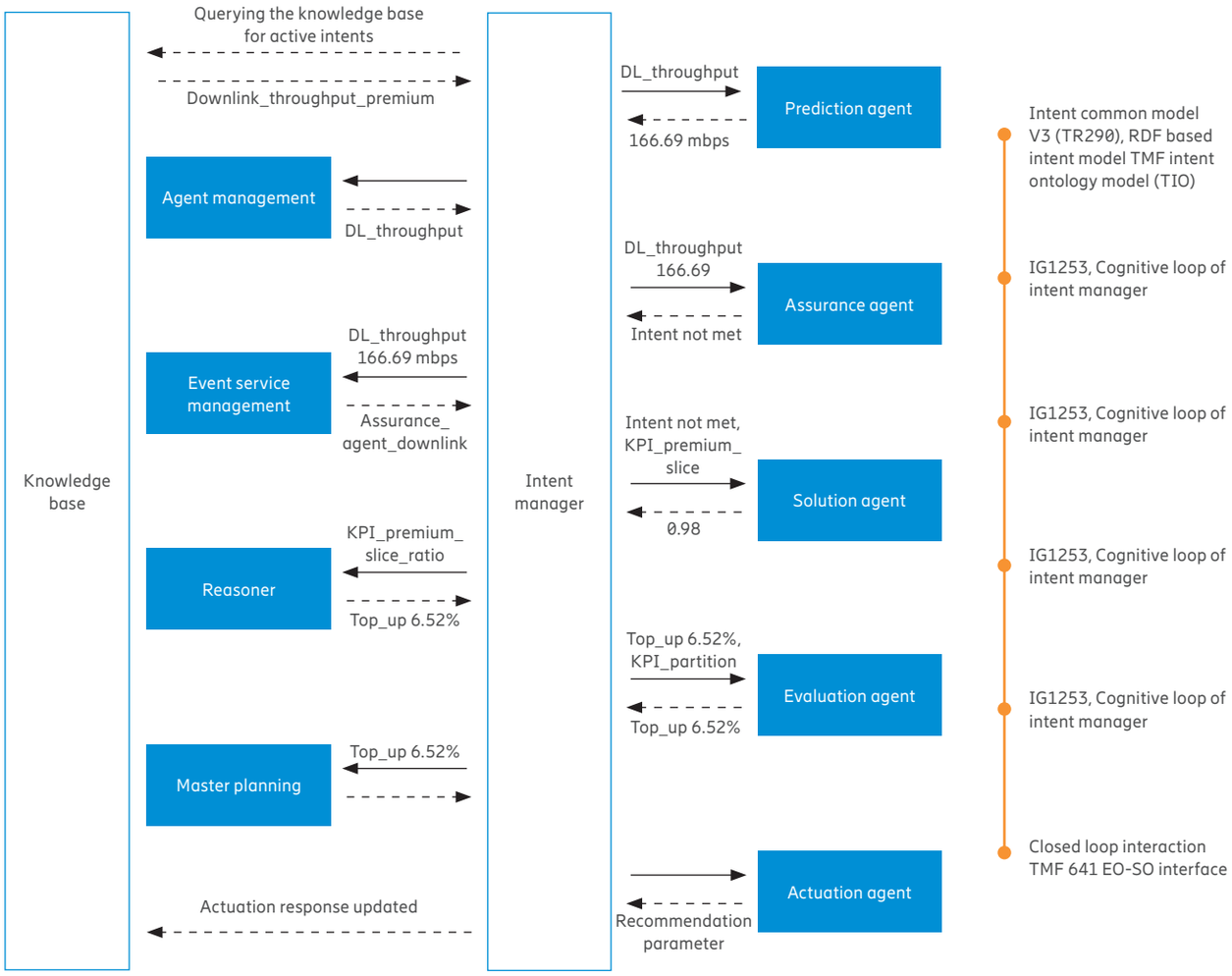
Intent handling encompasses "prediction and detection, on the basis of registered intents." These are crucial for understanding and presenting the current and future state of the network. The primary objective is to identify whether the owner's intent will be met and to detect any violations. When it becomes clear that the intent cannot be achieved based on the network's observed current – and predicted future – state, the secondary objective involves dynamically pinpointing potential solutions. This is another essential facet of intent handling in AI intent-based operations.

Solution formulation

When an intent is not met, the intent management function will engage solution proposal reasoners to dynamically come up with solutions to meet the intent. The proposals could be a combination of preexisting strategies, or an exploration of new solutions, based on machine reasoning. Once the solutions have been proposed, an evaluation agent performs a solution evaluation to determine which solution will be the most effective. Utility is an important aspect of intent, as it provides information about the properties of a preferable outcome and enables self-reflective operation for autonomous adaptation of the system's behavior. Once a solution is selected, the actuation orchestration is invoked to actuate the solution in the network. The visibility or the grounding agents will continue to provide current and predicated network states and – if the solution solved the issue – the intent will be met and the system resumes the intent-handling loop.

The Ericsson Operations Engine (EOE) journey from manual operation to operation with automated execution saw clear advantages, such as lead time of fulfilment after a service order and 24/7 attention of the system with immediate reaction. These have significantly improved service quality and decreased the operational costs to the customers.

Figure 5: Cognitive loop of service intent manager for 5G network slicing



Observing tangible benefits through four different premium SLAs:

- **100 percent slice user throughput SLA achieved, vs. 70 percent achievement without IBO.**
  Due to predictions being provided well in advance and dynamic changes on the go, the solution helps to achieve 100 percent SLA.
- **12 conflicting premium service SLA breaches avoided per cell, per day.**
  This means that the machine decides on actions to resolve conflicts in an order of priority stipulated by the customer.
- **8–10 percent better throughput of premium services vs. lower throughput and intent violation for premium services without IBO.**
  Overall, the machine is helping to achieve a better experience for VIP/Premium customers in a fully autonomous manner.
- **15–17 hours partition overuse prediction shared with customer per week/partition vs. no overuse visibility to customer.**
  Customers can monetize the overuse and agree required commercial agreements with other service providers.

- **Bandwidth overuse**
  Protection from bandwidth overuse by detecting such scenarios in real time
- **Intent handling**
  Intelligent solution to handle conflict and decide priority for bandwidth top-up
- **Guaranteed SLA**
  Guaranteed SLA by dynamic top-up decision as per KPI prediction and monitoring slice
- **5G monetization**
  Early higher partition share percentage bundle

Summary

Intent-based and autonomous networks will play a pivotal role in advancing network infrastructure by simplifying service definition and network operation, particularly within the dynamic 5G environment. While the objective of AI intent-based operations is to efficiently manage the complexities and variations of services, their implementation is just one part of the autonomous journey, leading toward a simplified and closed-loop environment allowing for almost zero-touch or fully autonomous operations of services.

The opportunities that AI intent-based operations could offer to service providers reach beyond just the financial. However, building service providers' confidence in autonomous network operations will be a gradual process and one of the first steps along this journey will be to establish a solid, intent-based foundation.

For further information and to learn more about AI intent-based operations, visit: www.ericsson.com/en/managed-services/ai-intent-based-operations

Contributors

Akshat Agarwal
Akshat is a Portfolio Director for Managed Network Services at Business Area Cloud, Software and Services with over 14 years of experience working in Network Operations, Automation and AI. He is currently working as Strategic Product Manager for Intent-based operations focusing on data-driven AI and enabling solutions to drive the autonomous journey.

Vivek Gautam
Vivek works at Head of Operations Assurance at Ericsson Malaysia and has over 16 years of experience working in Core & Cloud Network Operations, Customer Support, Managed Services and Automation. As an expert in Automated Network Operations, his current role is leading operations assurance for the world's first "six-operator wholesale 5G network", in partnership with Digital Nasional Berhad (DNB) Malaysia.

About Ericsson

Ericsson enables communications service providers and enterprises to capture the full value of connectivity. The company's portfolio spans the following business areas: Networks, Cloud Software and Services, Enterprise Wireless Solutions, Global Communications Platform, and Technologies and New Businesses. It is designed to help our customers go digital, increase efficiency and find new revenue streams. Ericsson's innovation investments have delivered the benefits of mobility and mobile broadband to billions of people globally. Ericsson stock is listed on Nasdaq Stockholm and on Nasdaq New York.

www.ericsson.com

Ericsson
SE-164 80 Stockholm, Sweden
Telephone +46 10 719 0000
www.ericsson.com

BCSS-24:019946 Uen