Intent AI Mediator (Mintent) for High-satisfaction-level Services

Kazuhisa Yamagishi, Masahiro Kobayashi, Shingo Horiuchi, and Kenichi Tayama

Abstract

This article introduces the intent AI (artificial intelligence) mediator called Mintent that coordinates and cooperatively controls networks, cloud servers, and applications on the basis of the respective requirements (intents) of service providers and users. This technology can be used to provide services on the basis of such intents.

Keywords: intent, Mintent, cooperative control

1. Introduction

Advances in communication network, cloud-server, and application technologies have made it possible to provide a higher satisfaction level of communication services. However, the resources of networks, cloud servers, and user terminals are limited, so the increase in the number of service users can cause network congestion and server resource constraints. When these situations occur, the throughput degrades and latency increases, and users cannot receive data properly at their terminals. As a result, users cannot experience a high satisfaction level of service. To avoid these quality degradations, service providers are deploying network, cloud-server, and application control technologies to overcome temporary resource constraints. However, improvements in these single-domain control techniques are limited. Therefore, it is important to coordinate and cooperatively control network, cloud-server, and application information.

The requirements (i.e., intents) that service providers and users demand of services generally differ for each service type. For example, a video-streaming service provider monitors the number of video users and the viewing time to encourage continuous use of the service. Thus, the service provider has an intent, i.e., a demand to increase the number of viewings or lengthen the viewing time. Similarly, the service provider of a connected vehicle has an intent to deliver video to the monitoring center without interruption to monitor what is happening inside and outside the vehicle. However, since it is not possible to provide services with an excessive amount of network, cloud-server, and application resources, these intents are converted into resource information for networks, cloud servers, and applications, which are coordinated to provide appropriate quality of service (QoS). Therefore, networks, cloud servers, and applications need to be controlled in a coordinated manner to provide appropriate QoS.

The intent artificial intelligence (AI) mediator called Mintent is introduced as a technology that coordinates and cooperatively controls networks, cloud servers, and applications on the basis of intents of service providers and users, as shown in **Fig. 1**.

2. Recent activities regarding intent-related technology in standardization

The Industry Specification Group on Experiential Networked Intelligence (ENI) of the European Telecommunications Standards Institute is developing a method for specifying the goals and requirements of operations that operators want to achieve as intents, which are described in restricted natural language. ENI 008 Intent Aware Network Automaticity released in March 2021 defines the architecture and functional blocks required to process an intent and discusses the

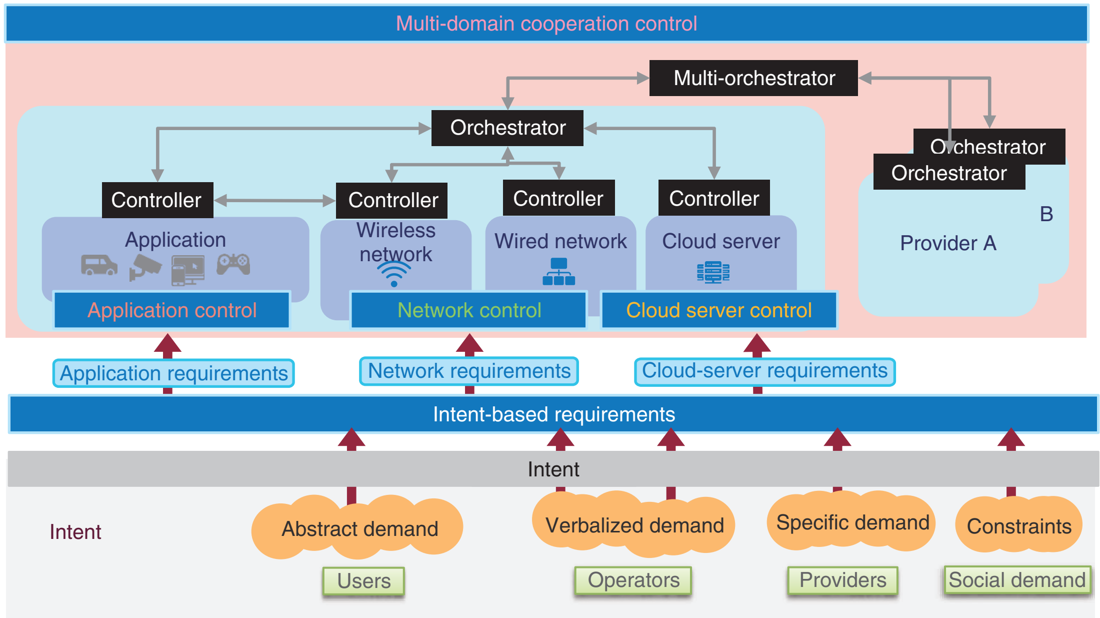


Fig. 1. Concept of intent AI mediator (Mintent).

lifecycle management of intent operations and elements such as external interface requirements. Proofs of concept applying intent management methods to the cloud and wireless networks are also being actively carried out.

TM Forum, an industry association for communication service providers and their suppliers, advocates a closed loop architecture to achieve autonomous network operations by taking into account the business, service, and network-resource layers, and the interworking of loops between each of these layers is defined as shown in **Fig. 2**. In this architecture, the autonomous network is achieved using the intents defined by the users and operators as the goal. This architecture has been discussed to study how to minimize the total operation cost.

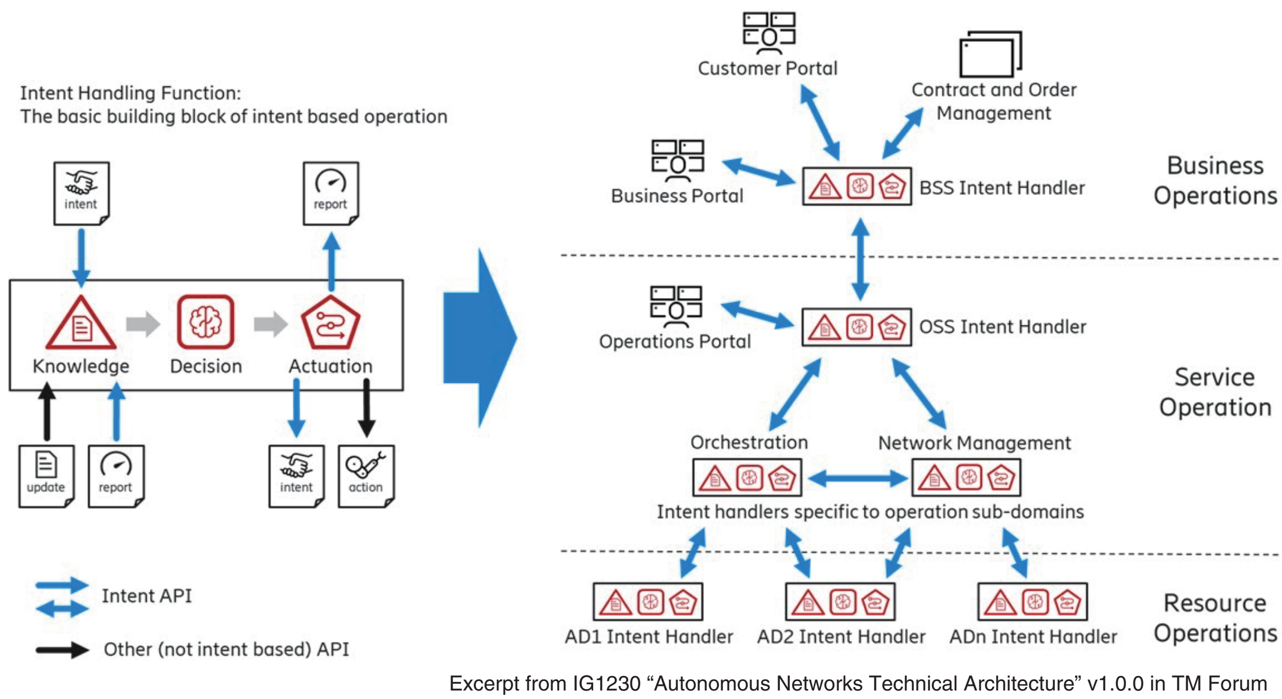
3. Mintent

To provide services that satisfy intents of service providers and users, it is not enough just to use respective control technologies in the network, cloud-server, and application domains. Therefore, Mintent was developed as a technology for sharing and coordinating information obtained from each domain to provide a higher satisfaction level of services by taking into account the intents of service providers and users (Fig. 1). An overview of the elemental technologies of Mintent is described below.

3.1 Application control technologies

Application control technologies to obtain video data with different video coding bitrates (i.e., adaptive bitrate streaming) or control video coding bitrates (i.e., Google Congestion Control) on the basis of network throughput, packet loss, and/or fluctuations in the amount of received data have become common for video streaming and web conferencing. These technologies provide a service without stalling by requesting video data with a lower coding bitrate when QoS (i.e., throughput, packet loss, and delay) degrades. However, such technologies cannot control on the basis of the service provider's or user's intent because the control is based only on QoS.

For example, service providers have an intent to avoid increasing content-delivery-network and network costs by providing services with excessive quality if they can maintain a certain level of suitable quality. Similarly, users have an intent to avoid excessive data consumption by receiving high-quality video. To satisfy these intents, we proposed technologies that do not transmit excessive data while maintaining a suitable quality level [1, 2]. These technologies consist of a quality-estimation model that estimates quality of experience (QoE) and request video data so that the QoE satisfies the suitable quality



AD: autonomous domain
API: application programming interface
BSS: business support system

Fig. 2. Autonomous network architecture using intent in TM Forum.

level. This prevents excessive quality and excessive data transmission, thus satisfying the intents of service providers and users.

3.2 Cloud-server control technology

As the environment surrounding the service changes, to maintain optimal resource allocation while satisfying QoE requirements, server resources need to be designed on the basis of various situations and requirements. With the diversification of services and increasing complexity of resource control, however, resource design and control based on the experience of service engineers or system performance-based methods are no longer adequate. Such a conventional approach has issues, such as service engineers are required to have much experience and high skill levels, the difficulty in providing services in accordance with user quality requirements and usage conditions, and the need to allocate extra resources to satisfy quality requirements.

To address these issues, we proposed a cloud-service-resource optimal-control technology that proactively and automatically calculates the amount of resources [3]. This technology handles user-experience quality requirements and service-provider performance requirements as intents, then the optimal cloud resource that satisfies the intent is calculated on the basis of the system environment, usage status, and planned usage information. The optimal cloud resource that satisfies the intent can be proactively controlled using an AI model that can predict performance on the basis of the system environment, usage status, and planned usage information.

3.3 Network control technology

Network control technology optimizes communication paths to maximize the efficiency of network equipment utilization while satisfying QoS requirements on the basis of the intents of users and network operators.

With the advancements in virtualization technologies such as software-defined networking and network functions virtualization, networks can be dynamically and flexibly controlled by software operations. As the services and terminals used become more diverse regarding user intent, QoS expected for the network is also becoming more diverse. In addition, network operators have an intent to use equipment efficiently and reduce equipment expansion.

To satisfy the intents of users and network operators, we proposed an optimal-route-allocation method to accommodate as many users as possible while satisfying QoS requirements calculated on the basis of users' QoE requirements by using network virtualization technology [4]. Since the optimum route allocation at a certain point in time is not always the optimum allocation in the future, this method improves the utilization efficiency of equipment by optimizing the resource allocation in consideration of future demand.

4. Multi-domain cooperative control technology

Multi-domain cooperative control coordinates the control for each domain (e.g., application, network, cloud server, and resource of other operators) to optimize the entire resource across domains while satisfying QoS requirements on the basis of the intent.

In future networks, such as IOWN (Innovative Optical and Wireless Network) and 6G (6th-generation mobile communication systems), controllers and orchestrators will be deployed in each domain and control the resources of the domain autonomously on the basis of the elemental technologies mentioned above. In this environment, individual optimization within a domain does not necessarily translate into optimization across domains. Therefore, the end-to-end intent may not be satisfied and resource utilization may be inefficient.

For optimal control across domains, we are developing *sharing of communication quality among domains* and *optimization of control order considering control execution time* as mechanisms to coordinate autonomous control by controllers in each domain. When communication quality in a certain domain degrades due to resource congestion or other reasons, resources will be over-allocated in other domains if the same amount of resources as in the normal case is reserved. Therefore, *sharing of communication quality among domains* prevents over-allocation of resources in other domains by sharing the communication quality achieved in each domain. As the types of control in each domain become more diverse, the execution time also varies. For example, rerouting in a network domain may take less than 1 minute, and cloud-server resource expansion may take 5–10 minutes. When congestion avoidance control is executed in multiple domains during resource congestion, the end-to-end communication quality deteriorates until the control with the longest execution time is completed. Therefore, *optimization of control order considering control execution time* shortens communication-quality deterioration by optimizing the content and order of control in consideration of the difference in execution time for each control.

5. Use cases of Mintent

Mintent enables multi-domain cooperative control such as the cooperation between wireless networks and applications, wired networks and applications, cloud servers and applications, and wired networks and cloud servers. The following sections describe the use cases and benefits of the coordination between wireless networks and applications and between wired networks and cloud servers.

5.1 Monitoring of in-vehicle camera videos

Studies for automated driving are rapidly advancing. However, fully automated driving will not be achieved immediately, and for the time being, driver assistance is the initial target. Specifically, images captured with an in-vehicle camera are sent to a monitoring center, where an observer monitors the images to assist the driver. In this case, the service provider has an intent to enable the non-stop delivery of surveillance video. Since wireless-network quality generally fluctuates, a high video-coding rate relative to the throughput causes packet loss, which distorts the video and prevents proper video monitoring. Therefore, current technologies either deliver video at a fixed coding rate designed to be safe with respect to wireless-network quality or deliver video at a low coding rate that does not reliably cause packet loss by capturing dynamic changes in the past wireless-network quality. Therefore, surveillance video with sufficient video quality cannot be provided to the observer, and there is a concern that the observer may fail to detect hazards both inside and outside the vehicle. Therefore, as shown in **Fig. 3**, Mintent dynamically controls the encoding rate by sharing the throughput value predicted as the future wireless-network quality with the encoder that encodes the in-vehicle camera video. This enables video transmission in accordance with the wireless-network quality and aims to transmit high-quality video in accordance with the wireless-network quality.

5.2 Robot control in smart factories

Automation of robot control is being studied for

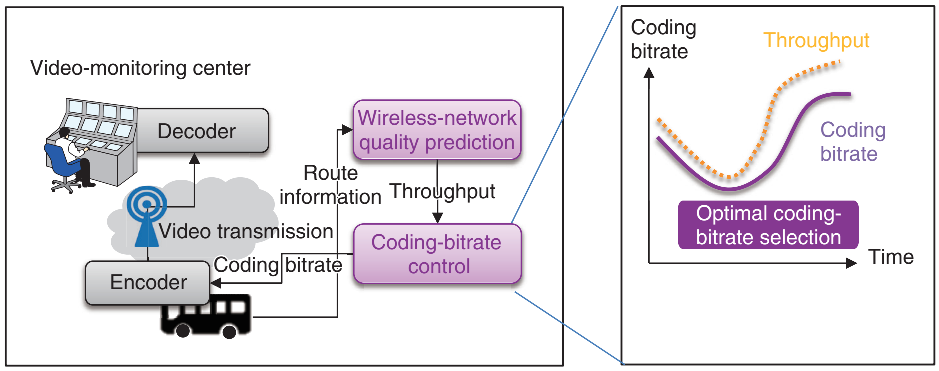


Fig. 3. Monitoring of in-vehicle camera videos.

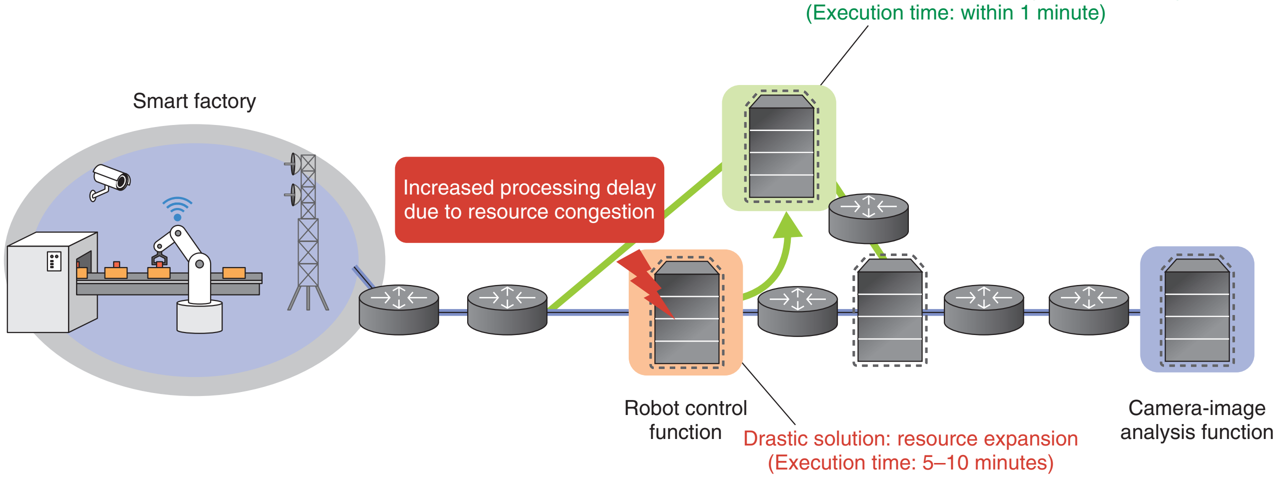


Fig. 4. Robot control in smart factory.

smart factories. Factory managers have an intent to keep robots running without stopping. To achieve this intent, communication-quality degradation that affects robot control must be minimized. Mintent coordinates the control of multiple domains and optimizes the control content and sequence by taking into account the control-execution time to shorten the quality-degradation period.

Consider the environment with a network domain and cloud-sever domain shown in **Fig. 4**. The robot-control function is deployed on the server close to the factory, and the camera-image analysis function is deployed on the server far from the factory. Consider the case in which processing delays increase due to temporary resource congestion on the server where the robot-control function is deployed. As a fundamental solution, the cloud-server-domain control needs to expand the resources of the server. However, since server-resource expansion takes about 5–10 minutes to execute, the processing delay continues to increase during this period. As a result, robot control is affected, and the factory manager's intent is not satisfied. Mintent can shorten the period of quality degradation by distributing the load to servers with sufficient resources by changing routes in network domains with control times of 1 minute or less as a temporary solution until the control of resource enhancement is completed.

6. Conclusion

In this article, the need for an intent AI mediator

(Mintent) was described. Subsequent articles will introduce the latest research and development of individual technologies at NTT laboratories.

References

[1] T. Kimura, T. Kimura, A. Matsumoto, and K. Yamagishi, "Balancing Quality of Experience and Traffic Volume in Adaptive Bitrate Streaming," IEEE Access, Vol. 9, pp. 15530–15547, 2021.

[2] M. Yokota and K. Yamagishi, "Quality-based Video Bitrate Control for WebRTC-based Teleconference Services," Electronic Imaging 2022, IQSP-333, Jan. 2022.

[3] C. Wu, S. Horiuchi, K. Murase, H. Kikushima, and K. Tayama, "Intent-driven Cloud Resource Design Framework to Meet Cloud Performance Requirements and Its Application to a Cloud-sensor System," J. Cloud Comp., Vol. 10, 30, 2021.

[4] K. Kawakami, M. Kobayashi, K. Yogo, K. Okuda, M. Sekiguchi, T. Kurahashi, S. Arai, T. Tsuchiya, and N. Shirai, "Network Control Technology to Realize the NetroSphere Concept," NTT Technical Review, Vol. 14, No. 10, Oct. 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201610fa4.html

Kazuhisa Yamagishi

Group Leader, NTT Network Service Systems Laboratories.

He received a B.E. in electrical engineering from Tokyo University of Science in 2001 and M.E. and Ph.D. in electronics, information, and communication engineering from Waseda University, Tokyo, in 2003 and 2013. Since joining NTT in 2003, he has been engaged in the development of objective quality-estimation models for multi-media telecommunications. From 2010 to 2011, he was a visiting researcher at Arizona State University. He has been contributing to International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Study Group 12 (SG12) on QoS, QoE, and Performance since 2005. He has been a vice-chair of SG12 since 2022, vice-chair of Working Party 3 in SG12 since 2021, and a co-rapporteur of Question 13/12 since 2017.

Masahiro Kobayashi

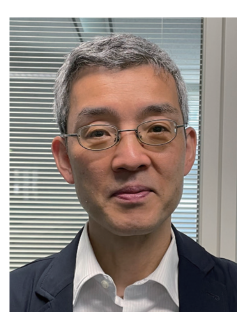
Senior Research Engineer, NTT Network Service Systems Laboratories.

He received a B.S. and M.S. in information science from Tohoku University, Miyagi, in 2007 and 2009. He joined NTT in 2009 and is currently studying network analysis, design, and control. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Shingo Horiuchi

Senior Research Engineer, NTT Access Network Service Systems Laboratories.

He received a B.E. and M.E. in engineering from the University of Tokyo in 1999 and 2001. He joined NTT Access Network Service Systems Laboratories in 2001. He has been researching and developing access network operation systems. He has also been involved in standardization efforts for operations support systems in TM Forum. He is a member of IEICE.



Kenichi Tayama

Group Leader, Senior Research Engineer, Supervisor, NTT Access Network Service Systems Laboratories.

He received a B.S. and M.S. from The University of Electro-Communications, Tokyo, in 1993 and 1995. Since joining NTT in 1995, he has been engaged in the research and development of access network operations and the planning and development of internal information technology systems and network operations and maintenance.