INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

SG13-TD142-R1/WP3

STUDY GROUP 13

Original: English

Question(s): 2/13

Geneva, 25 November 2022

TD

Source: Editors

Title: Draft new Recommendation ITU-T Y.SAN: "Requirements and architecture of service aware network"

Contact:	Xinxin Yi China Unicom P.R.C	Tel: +86-10-68799999 E-mail: yixx3@chinaunicom.cn
Contact:	Ran Pang China Unicom P.R.C	Tel: +86-10-68799999 E-mail: pangran@chinaunicom.cn
Contact:	Qianying Zhao China Telecom P.R.C	Tel: + 86-10-50902878 E-mail: zhaoqy50@chinatelecom.cn
Contact:	Shuai Zhang China Unicom P.R.C	Tel: +86-10-68799999 E-mail: zhangs366@chinaunicom.cn

Abstract: This document is the output of draft Recommendation: "requirements and architecture of service aware network". This document includes the results of discussion in the Q2/SG13 meeting which was held on 14-25 November 2022.

This document is the output document for draft Recommendation ITU-T Y.SAN "requirements and architecture of service aware network" based on the contribution to initial a new draft Recommendation on requirements and architecture of service aware network, and the meeting discussion (14-25, November, 2022 SG13 RGM meeting).

Base text from the previous SG13 meeting (14-25, November, 2022).

C96-R1	China Unicom, China Telecommunications Corporation	Propose to initial a new draft Recommendation on requirements and architecture of service aware network	Accepted with modification

This contribution proposes to initial a new draft Recommendation on requirements and architecture of service aware network (tentative name: Y.SAN).

Meeting Result:

– Accepted with modification.

Draft new Recommendation ITU-T Y.SAN

Requirements and architecture of service aware network

1. Scope

This draft Recommendation provides the requirements and architecture of service aware network to enhance service awareness and end-to-end network scheduling capabilities.

2. References

The following ITU-T Recommendations and other references contain provisions, which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Y.2301] Recommendation Y.2301 (2013), *Network intelligence capability enhancement - Requirements and capabilities.*

[ITU-T Y.2303] Recommendation Y.2303 (2015), *Network intelligence capability enhancement - Awareness functional architecture.*

[ITU-T Y.2323] Recommendation Y.2323 (2018), *Requirements and capabilities of orchestration in next generation network evolution.*

[ITU-T Y.2324] Recommendation Y.2324 (2019), *Functional architecture of orchestration in next generation network evolution (NGNe).*

[Editor's Note] The ongoing item ITU-T Y.NGNe-MC-reqts、ITU-T Y.NGNe-NCI-reqts focus on NGN evolution for support network and cloud interworking, may be concerned in the future.

3. Definitions

3.1 Terms defined elsewhere

[To be continued]

3.2 Terms defined in this Recommendation

[To be continued]

4. Abbreviations and acronyms

[To be continued]

5. Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option, and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with this Recommendation.

6. Background and motivations

With the construction of 5G network and the evolution of cloud era, the rapid development of emerging applications such as online games, live video and video conference have put forward new performance and functional requirements for terminals, OTT service providers and operators' networks, including low latency, low jitter, high reliability and large bandwidth and other differentiated indicators.

At present, the end-to-end network capability is not enough to fully match the personalized service demand, or to achieve fine and accurate service capability matching.

The continuous changes in service requirements are pushing the network capabilities of 5G to support the real-time awareness of user service requirement, on-demand scheduling of services and resources, full opening of network service capability.

Therefore, it is necessary to build a service aware network with the guidance of service awareness, service scheduling and service differentiation which can aware users, service requirements, resource and service status and so on in real time and provide end-to-end network scheduling and differentiated service provision.

7. Scenarios

7.1 Scenario of service scheduling based on application requirement awareness

As shown in the figure, on the access side, the CPE can be aware of user information, service information, and network requirements. On the cloud side, the service gateway can be aware of service deployment status and pool resources. Based on the awareness of users, services and resources, operator provides a set of network orchestration management and control system, to accomplish end-to-end connection, intelligent scheduling and differentiated services for user1 and user2.

User1, an advanced user, a delay-sensitive user, needs a service to access function1 and function 2, of which the end-to-end latency is less than 10ms and the bandwidth is 100 mbit/s. The service need to select a low load rate resource service node because a lower load rate resource pool can meet the service quality requirements better.

User2, a mid-level user, needs a service to access function 2, of which the end-to-end latency is less than 30ms and the bandwidth is 50 mbit/s. The service has no requirements for selecting resource pools.

By the awareness of user, services and resources in the network in real time, mapping user requirements to the carrier network, operators may provide delay-sensitive users with low-latency paths through network slices, or establish a path to the resource pool with low load rate for delay-sensitive users based on the resource pool load.

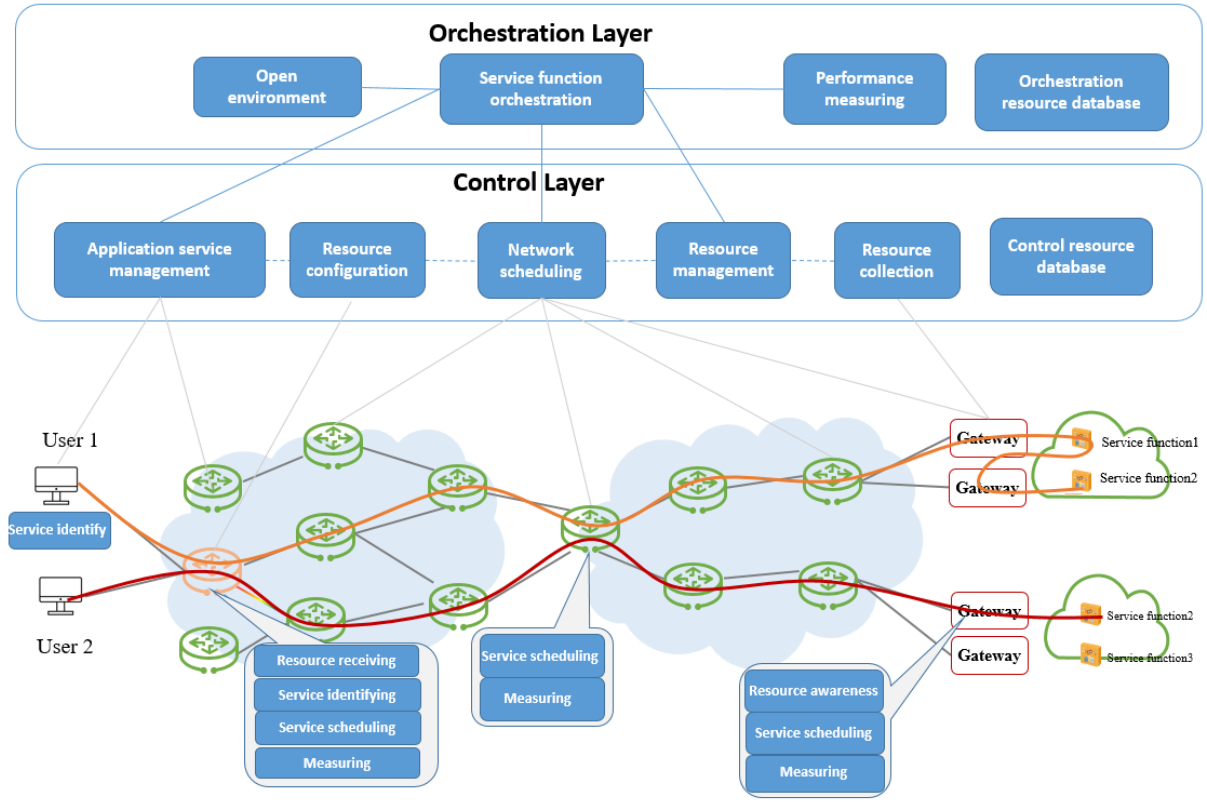

Figure 7-1 scenario of service scheduling based on application requirement awareness

7.2 Scenario of real-time awareness on the move

Cloud game deploys game applications in data centers, and realizes the functions including the logical process of game command control, video rendering and other tasks that have high requirements for chips, and the tasks of game acceleration. With the distributed deployment, each function of cloud game may be deployed on the edge data center near the user side on demand. The edge data center sends the game video stream information to the terminal, and receives the user's control instruction information for processing. Users can make corresponding operation instructions according to the received video stream information, and get quick response.

In mobility scenarios, there are new requirements for rapid awareness, adjustment and scheduling of paths and resource pools in a unified manner for wireless, carrier network, and edge resource nodes. From the end-to-end perspective, we need to consider how to be accurately aware of users, how to solve engagement problems for mobile users, how to solve the problem of resource pool adjustment in different access locations.

By identifying user information, application information and resource information, users can choose appropriate nodes which provide game acceleration function, and operators provide the specific network forwarding services, which can ensure deterministic latency of multi-party network of multiple players for better game experience.

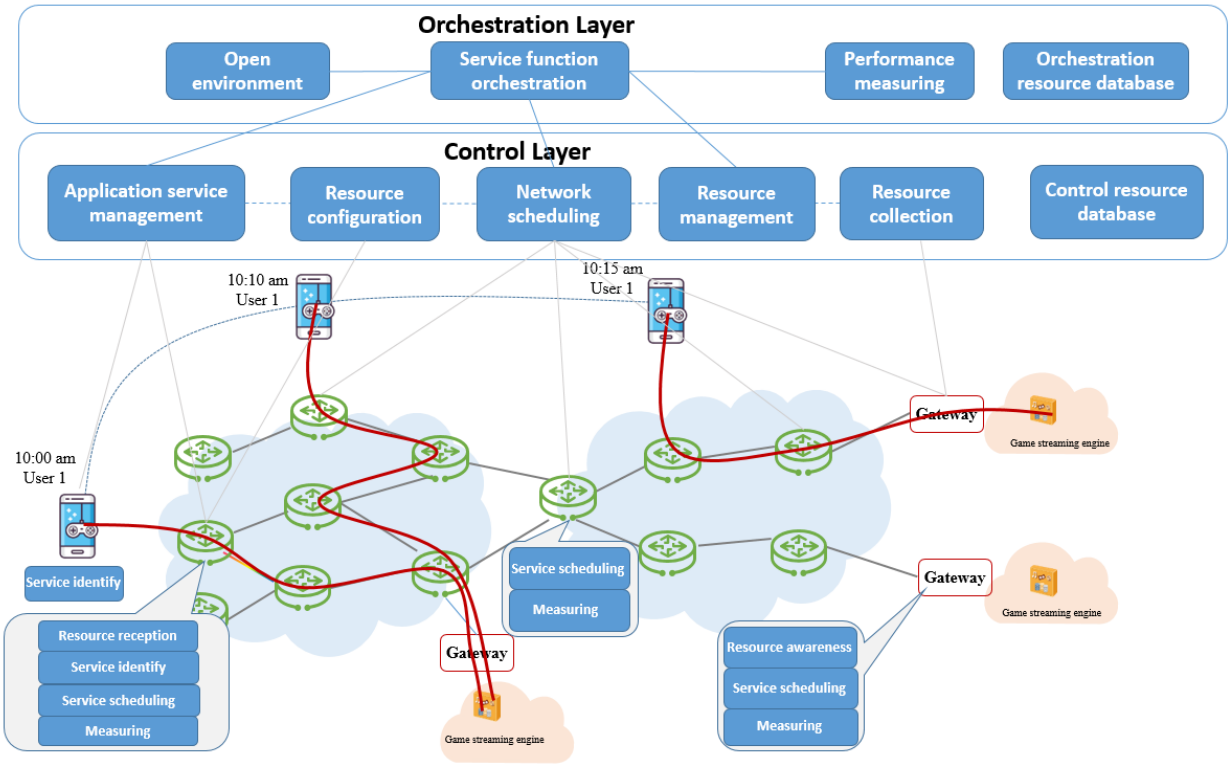

Figure7-2 scenario of real-time awareness on the move

As shown in the figure, when a user takes a train from city A to city B, during the journey, the user accesses a cloud game. At the access side, the user accesses the carrier network from the nearest base station, and then selects the appropriate node to access the game according to the game streaming engine deployment and resource pool situation around.

Since the user is on the move, to ensure the game experience, that is, to ensure the continuity of the service, network needs to be aware of user information and application information, resource pool load rate and so on in real time, and switch to another appropriate resource pool based on the above information. In the process of switching, the current game state should be guaranteed to provide a good game experience.

8. Functional architecture of service awareness network

[Editor's Note] This clause will give the general framework of service aware network.

The service awareness will help improve end-to-end service quality, provide precise, flexible and differentiated service capabilities for different users, and satisfy new productization possibilities for the opening of network capabilities of operators.

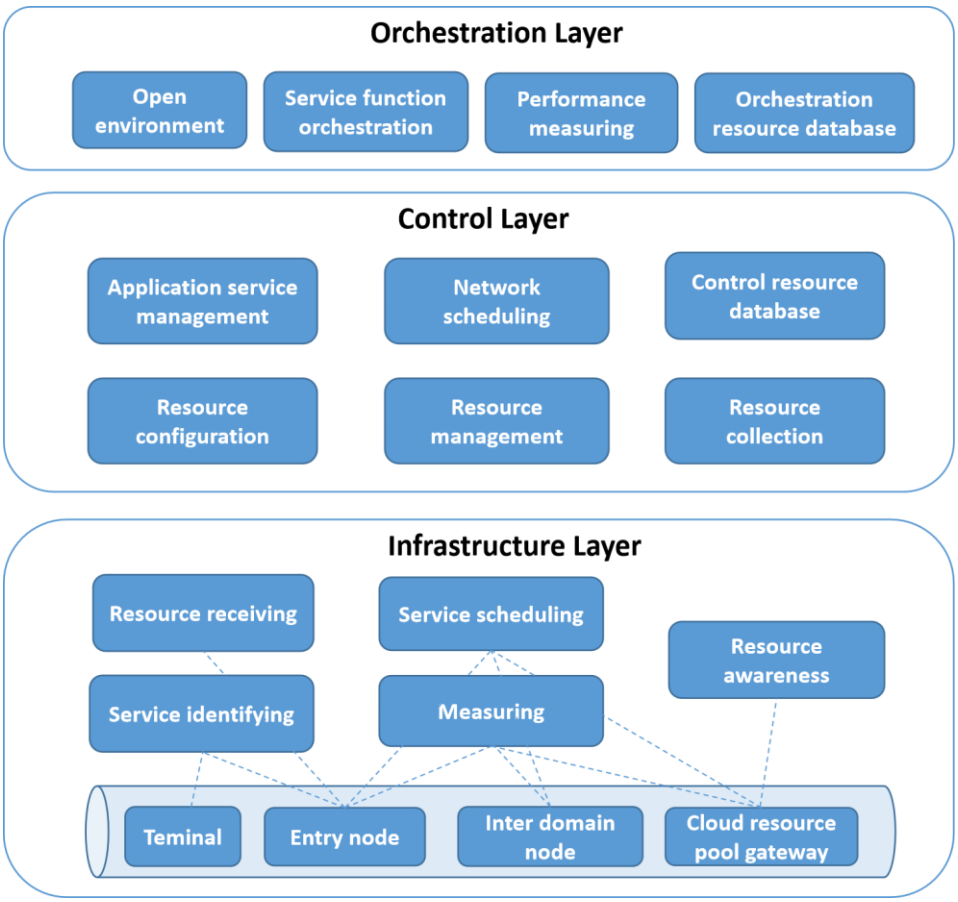

Figure8-1 Functional architecture of service aware network

Figure 8-1 shows the overall architecture of service aware network, which includes the infrastructure layer, the control layer, and the orchestration layer.

8.1 Infrastructure Layer

[Editor's Note] This clause will give the description of Infrastructure Layer of service awareness network.

The infrastructure layer includes network forwarding equipment and services functions. As shown in Figure 8-1, the infrastructure includes terminal, entry node, inter domain node, cloud resource pool gateway.

The terminal refers to the user terminal equipment, it needs support service identifying module.

The entry node refers to the entry location of the network, and the starting location of the service in the network, it needs to include resource receiving module, service identifying module, service scheduling module, and measuring module.

The service identify module can be completed either at the terminal or at the entry node.

The inter domain node refers to the node interconnected among multiple network domains, it needs to include service scheduling module, and measuring module.

The cloud resource pool gateway refers to the gateway node connecting with the cloud resource pool where services and resources are located, it needs support service scheduling module, measuring module and resource awareness module.

The above four types of nodes need new functions in this architecture. Other nodes in the infrastructure layer, such as intermediate forwarding nodes, only need to support existing functions.

The service identifying module is used to identify the service identifier carried by the user request.

The resource receiving module of the entry node is used to receive and record the network information and service information, those information is obtained and stored by the upper control layer in real time, customized for each entry node and announced to entry node. Only the entry node needs this function module.
The service scheduling module performs service scheduling and routing decisions based on the received network information. The network entry node need this module, and some inter domain nodes or cloud resource pool gateway also need to service scheduling.

The measuring module is used to complete the measurement of service performance indicators. Generally, this module is required for all network nodes requiring performance measurement.

The resource awareness is used to aware and obtain resource information when these information reported through the cloud pool gateway. If those information reported through cloud or resource manager platform, then the resource awareness module is unnecessary to the cloud pool gateway.

8.2 Control Layer

[Editor's Note] This clause will give the description of Control Layer of service aware network.

The control layer includes application service management, resource configuration, network scheduling, resource management, resource collection, and control layer database;

The application service management module is used to classify, identify and manage service applications;

The resource collection module is used to obtain and store the information of each node in the network in real time;

The resource management module processes and analyses the collected data, and is used to formulate delivery information for each entry node in the network;

The resource configuration module sends the data configuration to each entry node based on the processing result of the management module, so that each entry node receives and stores the corresponding network information;

The network orchestration module makes routing decisions based on the nodes and network node information in the network in a centralized way for non-delay sensitive services;

The control layer database is used to store and record the network, and business information of the management and control layer data.

8.3 Orchestration Layer

[Editor's Note] This clause will give the description of Orchestration Layer of service aware network.

The orchestration layer includes open environment, service function orchestration, performance measurement, and orchestration layer database modules.

The open environment is used to support the requirements of openness;

The service function orchestration is used to achieve the selection and orchestration of service resources;

The performance measurement is used to realize the real-time monitoring of service quality and fault location;

The orchestration layer database stores and records the network, business and other data information of the orchestration layer.

9. Requirements of service awareness network

[Editor's Note] This clause will provide the requirement of service aware network.

10. Service Process

[Editor's Note] This clause will provide the requirement of service process.

The resource collection module of the control layer obtains the service information in the cloud pool in the network in real time, and stores it in the corresponding database;

At the same time, the control layer also collects and obtains the network node information in the network in real time, and stores it in the corresponding database.

Then the resource management module processes the service information, and also the data based on different dimensions such as service, region, time or user priority, formulates corresponding delivery information for each entry node, and then notifies the entry node through the resource configuration module.

At the same time, the control layer processes the information and network information and reports it to the orchestration layer. The orchestration layer selects service resources and arranges services based on user needs.The orchestrated path is delivered to the entry node.

In this way, the entry node receives and stores the information sent by the upper-layer system, and the entry node can implement real-time routing decisions or direct tunnel iteration according to the specific needs of the business. When user 1 is an advanced user with delay-sensitive services, the entry node makes real-time routing decisions based on the delivered information; when user 2 is an intermediate user with non-delay-sensitive services, the entry node makes forwarding decisions based on the delivered information that has been calculated and completed. The user identification and user demand awareness of user 1 and user 2 are completed by the service identification module of the user terminal or entry node.

When the state of the service involved in user 1 changes, or the network state changes, or the service demand changes, the entry node can quickly complete the re-selection of the service and the recalculation of the route based on the information sent, so as to achieve rapid convergence effect.

Based on the above functions, the entry node completes service awareness, business differentiation strategies and routing decisions according to different business categories and needs, so as to meet the intelligent scheduling and connection services of different businesses.

11. Security considerations

[Editor's Note] This clause will provide security considerations.

_____

Annex 1
A.1 justification for proposed draft new Recommendation Y.SAN

Question:	2/13	**Proposed new ITU-T Recommendation**	Geneva, 25 November 2022	
Reference and title:	Draft Recommendation ITU-T Y.SAN "Requirements and architecture of service aware network"			
Base text:	TD142/WP3		**Timing:**	November, 2024
Editor(s):	Xinxin Yi , China Unicom, China E-mail: yixx3@chinaunicom.cn Ran Pang , China Unicom, China E-mail: pangran@chinaunicom.cn Qianying Zhao, China Telecom, China E-mail: zhaoqy50@chinatelecom.cn Shuai Zhang , China Unicom, China E-mail: zhangs366@chinaunicom.cn		**Approval process:**	AAP

Scope (defines the intent or object of the Recommendation and the aspects covered, thereby indicating the limits of its applicability):

This draft Recommendation provides the requirements and architecture of service aware network to enhance service awareness and end-to-end network scheduling capabilities.

Summary (provides a brief overview of the purpose and contents of the Recommendation, thus permitting readers to judge its usefulness for their work):

The rapid development of emerging applications such as online games, live video and video conference have put forward new performance and functional requirements for terminals, OTT service providers and operators' networks, including low latency, low jitter, high reliability and large bandwidth and other differentiated indicators.

At present, the end-to-end network capability is not enough to fully match the personalized service demand, or to achieve fine and accurate service capability matching.

The continuous changes in service requirements are pushing the network capabilities to support the real-time awareness of user service requirement, on-demand scheduling of services and resources, full opening of network service capability.

Therefore, it is necessary to build a service aware network with the guidance of service awareness, service scheduling and service differentiation which can be aware of users, service requirements, resource and service status and so on in real time and provide end-to-end network scheduling and differentiated service provision.

Relations to ITU-T Recommendations or to other standards (approved or under development):

ITU-T Y.2301、ITU-T Y.2303、ITU-T Y.2323、ITU-T Y.2324、ITU-T Y.NGNe-MC-reqts、ITU-T Y.NGNe-NCI-reqts、ITU-T Y.SFO

The Recommendation ITU-T Y.2301、ITU-T Y.2303、ITU-T Y.2323、ITU-T Y.2324 and ongoing draft Recommendation ITU-T Y.NGNe-MC-reqts、ITU-T Y.NGNe-NCI-reqts、ITU-T Y.SFO focuses on NGN evolution for support network and cloud interworking and enhancement for NGNs which support some intelligence capabilities for the provisioning of services.

This new work item focuses service aware network architecture to enhance service awareness and end-to-end network scheduling capabilities.

IETF Application-aware Networking (APN) Framework

The ongoing IETF draft "Application-aware Networking (APN) Framework" mainly describes a new framework, where application-aware information (APN attribute) including application-aware identification (APN ID) and application aware parameters (APN Parameters), is encapsulated at network edge devices and carried along with the encapsulation of the tunnel that is used by the packet to traverse the APN domain.

Identifying user information and application information on access side by APN technology is a single point of functional technology capability. This new work item focuses on service aware network architecture and service scheduling and differentiated service provision and so on.
Liaisons with other study groups or with other standards bodies: IETF
Supporting members that are committing to contributing actively to the work item: China Unicom, China Telecom

_____