ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T Q.4141 (12/2023)

SERIES Q: Switching and signalling, and associated measurements and tests

Protocols and signalling for computing power networks

Requirements and signalling of intelligence control for the border network gateway in computing power networks

ITU-T Q-SERIES RECOMMENDATIONS

Switching and signalling, and associated measurements and tests

SIGNALLING IN THE INTERNATIONAL MANUAL SERVICE	Q.1-Q.3
INTERNATIONAL AUTOMATIC AND SEMI-AUTOMATIC WORKING	Q.4-Q.59
FUNCTIONS AND INFORMATION FLOWS FOR SERVICES IN THE ISDN	Q.60-Q.99
CLAUSES APPLICABLE TO ITU-T STANDARD SYSTEMS	Q.100-Q.119
SPECIFICATIONS OF SIGNALLING SYSTEMS NO. 4, 5, 6, R1 AND R2	Q.120-Q.499
DIGITAL EXCHANGES	Q.500-Q.599
INTERWORKING OF SIGNALLING SYSTEMS	Q.600-Q.699
SPECIFICATIONS OF SIGNALLING SYSTEM NO. 7	Q.700-Q.799
Q3 INTERFACE	Q.800-Q.849
DIGITAL SUBSCRIBER SIGNALLING SYSTEM NO. 1	Q.850-Q.999
PUBLIC LAND MOBILE NETWORK	Q.1000-Q.1099
INTERWORKING WITH SATELLITE MOBILE SYSTEMS	Q.1100-Q.1199
INTELLIGENT NETWORK	Q.1200-Q.1699
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2000	Q.1700-Q.1799
SPECIFICATIONS OF SIGNALLING RELATED TO BEARER INDEPENDENT CALL CONTROL (BICC)	Q.1900-Q.1999
BROADBAND ISDN	Q.2000-Q.2999
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR THE NGN	Q.3000-Q.3709
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR SDN	Q.3710-Q.3899
TESTING SPECIFICATIONS	Q.3900-Q.4099
PROTOCOLS AND SIGNALLING FOR PEER-TO-PEER COMMUNICATIONS	Q.4100-Q.4139
PROTOCOLS AND SIGNALLING FOR COMPUTING POWER NETWORKS	**Q.4140-Q.4159**
PROTOCOLS AND SIGNALLING FOR QUANTUM KEY DISTRIBUTION NETWORKS	Q.4160-Q.4179
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2020	Q.5000-Q.5049
COMBATING COUNTERFEITING AND STOLEN ICT DEVICES	Q.5050-Q.5069

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Q.4141

Requirements and signalling of intelligence control for the border network gateway in computing power networks

Summary

Recommendation ITU-T Q.4141 provides a network solution to dynamically and flexibly schedule computing tasks at the border network gateway (BNG) based on real-time computing resource performance, network performance, cost and other multi-dimensional factors related to business needs, thereby improving resource utilization, network utilization efficiency, and business user experience.

This Recommendation aims to study the requirements and signalling of intelligence control for the border network gateway in a computing power network (CPN).

History*

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T Q.4141	2023-12-14	11	11.1002/1000/15724

Keywords

Computing power network, CPN gateway, intelligence control.

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Q.4141

Requirements and signalling of intelligence control for the border network gateway in computing power networks

1 Scope

This Recommendation focus on the requirements architecture and signalling procedure of intelligence control for the border network gateway (BNG) devices in a computing power network (CPN). This Recommendation provides a network solution to dynamically and flexibly schedule computing tasks at the border network gateway based on real-time computing resource performance, network performance, cost and other multi-dimensional factors related to business needs, thereby improving resource utilization, network utilization efficiency, and user experience.

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 Q.3719] Recommendation ITU-T Q.3719 (2019), *Signalling requirements for the separation of control plane and user plane in a virtualized broadband network gateway (vBNG).*

[ITU-T Q.4140] Recommendation ITU-T Q.4140 (2023), *Signalling requirements for service deployment in computing power networks.*

[ITU-T Y.2501] Recommendation ITU-T Y.2501 (2021), *Computing power network – Framework and architecture.*

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 computing power network (CPN) gateway: A border gateway with the functions of service perception, computing power perception, and computing power routing, etc., that can dynamically select forwarding paths and service nodes based on the calculation of service requirements, computing power information and network information, to realize the global optimization of computing power and network resources.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

ASIC Application Specific Integrated Circuit

BGP	Border Gateway Protocol
BNG	Border Network Gateway
CPN	Computing Power Network
CPN-GW	Computing Power Network Gateway
CPN-GW-CP	Computing Power Network Gateway Control Plane
CPN-GW-UP	Computing Power Network Gateway User Plane
CPU	Central Processing Unit
DC	Data Centre
EVPN	Ethernet Virtual Private Network
GPU	Graphics Processing Unit
IGP	Interior Gateway Protocol
IoT	Internet of Things
IP	Internet Protocol
MPLS	Multiple Protocol Label Switch
NP	Network Process
NPU	Neural network Processing Unit
OAM	Operation Administration and Maintenance
PE	Provider Edge
RESTful	Representational State Transfer
SDN	Software-Defined Networking
SR	Segment Routing
SRv6	Segment Routing Internet Protocol Version 6
SSL	Secure Socket Layer
VPN	Virtual Private Network
VR/AR	Virtual Reality/Augmented Reality
VxLAN	Virtual extensible Local Area Network
WAN	Wide Area Network

5 Conventions

None.

6 Requirements

6.1 Overview

In a computing power network (CPN), the computing power network gateway (CPN-GW) is able to collect network resource and status information and computing power information, and so conduct the scheduling policy of computing power intelligently. The CPN gateway perceives computing resource information and network resource information, notifies the current computing resource status as routing information to the network, and routes computing tasks to appropriate computing nodes to achieve optimal overall system and user experience. Computing power awareness and computing power routing are the two core capabilities of the CPN gateway.

CPN gateway awareness is a comprehensive perception of the performance of computing resources, real-time load, network status, and business requirements. It mainly needs to know how many computing resources are in the network and what kinds of computing power are required by users. Computing power awareness includes computing resource perception, network resource perception, and business demand perception. Computing power routing is the process of integrating and distributing network and computing resource information, and achieving resource information announcement and information sharing across the entire network.

In a CPN, the intelligence control functions for computing power scheduling are deployed in the network control plane. As described in [ITU-T Q.3719], the architecture of the BNG is separated into a control plane and a data plane. It is therefore compatible to extend the scope of functions in the BNG control plane for intelligence control of computing power scheduling in CPN.

However, in spite of the advantages and compatibility of the BNG for scheduling computing power it remains a challenge to schedule computing power at the CPN gateway in a dynamical and flexible way considering both computing resources and network performances.

This clause focuses on the general requirements and specific requirements for the CPN gateway.

6.2 General requirements

6.2.1 Bandwidth requirement

It is required to support reserving the bandwidth resources in advance to transmit massive data when scheduling the computing power for those kinds of services which require a lot of bandwidth resources, such as virtual reality/augmented reality (VR/AR).

6.2.2 Latency requirement

It is required to support lower and deterministic end-to-end latency when scheduling the computing power for services which are sensitive to latency, such as tactile Internet.

6.2.3 Jitter requirement

It is required to support the lower and deterministic jitter of network latency when scheduling the computing power for those kinds of services which are sensitive to the jitter of network latency, such as telemedicine so that the jitter of network latency should be restricted within a certain range.

6.2.4 Packet loss rate requirement

It is required to support lower packet loss rate when scheduling the computing power for those kinds of services, such as smart factory, which are sensitive to packet loss rate.

6.2.5 Reliability requirement

It is required to support the reliability of communication when scheduling the computing powers for that kinds of services which require high reliability of communication, such as self-driving vehicles.

6.2.6 Security requirement

It is required to support higher security when scheduling the computing powers for that kinds of services which have particular requirements for security, such as smart home which requires the user's data to be processed locally.

6.3 Specific requirements for the CPN gateway

Besides the general functions of a traditional border gateway, a CPN gateway also needs to perceive computing resource information and network resource information, distribute the current computing resource status and network resource status to the network along with routing information, and route computing tasks to the appropriate computing nodes to achieve overall system optimization and user experience optimization. To this end, the CPN gateway needs to implement specific functions, such as computing resource awareness, network resource status awareness, service requirements awareness, computing power routing capability, computing power resource information distribution, and computing power routing table generation, etc.

6.3.1 Computing resource awareness

Computing resource awareness is the perception of computing resources of the computing power resource pools, including their performance, real-time load and other related information, which usually includes the IP address, storage capacity, etc. For example, cloud resource pools are usually centrally managed by the cloud management platforms. The CPN gateway can interact with the cloud management platform through interfaces such as representational state transfer (RESTful) APIs to obtain relevant computing power resource information about the cloud resource pools.

6.3.2 Network resource status awareness

The perception of network resource status information usually includes bandwidth, delay, jitter, packet loss rate, etc. Taking network delay as an example, since computing power resource pools are in geographically distributed locations, the network path from the user to the computing power resource pool will also change according to the network congestion status. Therefore, it is necessary to detect the network delay information between the user and each computing power resource pool.

6.3.3 Service requirements awareness

The CPN gateway at the user entrance receives service requests and perceives service requirements, including network requirements (such as delay, jitter, etc.) and computing power resource requirements (such as computing power type, computing power demand parameters, etc.), based on computing power measurement standards and matches the available computing power to a specific algorithm. It not only accurately matches the service requirements of specific applications, but also dynamically and real-time schedules the computing power network to optimize the computing resources and the network resources.

6.3.4 Computing power routing capability

Computing power routing introduces computing power resource information into the routing domain, and dynamically selects the "forwarding path + destination service node" that meets the service requirements by perceiving the user's service requirements, computing power resources, and network resources status, and then schedules the service along the designated path to service nodes to achieve global optimization of computing power resources and network resources.

6.3.5 Computing power resource information distribution

The CPN gateway will advertise computing power resource information to the adjacent CPN gateways, computing power routing nodes or the controller. The information notified includes computing power resource information such as processor's type, overload and memories, as well as network-related indicators in the computing power resource pools such as delay and jitter, etc.

6.3.6 Computing resource routing table generation

After obtaining the distribution status of computing power resource nodes, the CPN gateway will generate a computing power routing table within the domain. In addition to the network parameters (such as cost) maintained in the IP routing table, the computing power routing table also contains computing power resource information, thereby building a "network + computing" global routing table.

7 Architecture of intelligence control for the CPN gateway

For the purpose of intelligence control for the CPN gateway, the architecture is designed to meet these requirements:

– The CPN gateway should collect the information of computing power resources and network resources in the central cloud and the edge cloud. The computing power resources information consists of central processing unit (CPU utilization), memory utilization, storage utilization, etc. and the network resources information consists of bandwidth utilization, latency, etc.

– The design of architecture should consider supporting computing power cross-domain route information. The CPN inter-domain route information should be constructed on all intra-domain route information maintained by the control plane of the CPN gateway. Additionally, the inter-domain and intra-domain computing task scheduling policies should be made based on the corresponding resources information and route information.

– The CPN gateway should interact with components contained in the CPN service layer to accept the service demands and to report the computing power resources information and network resources information.

– The CPN gateway should be able to allocate the resources and schedule the network connection between the consumers and the providers.

The architecture of the CPN gateway is shown in Figure 7-1.

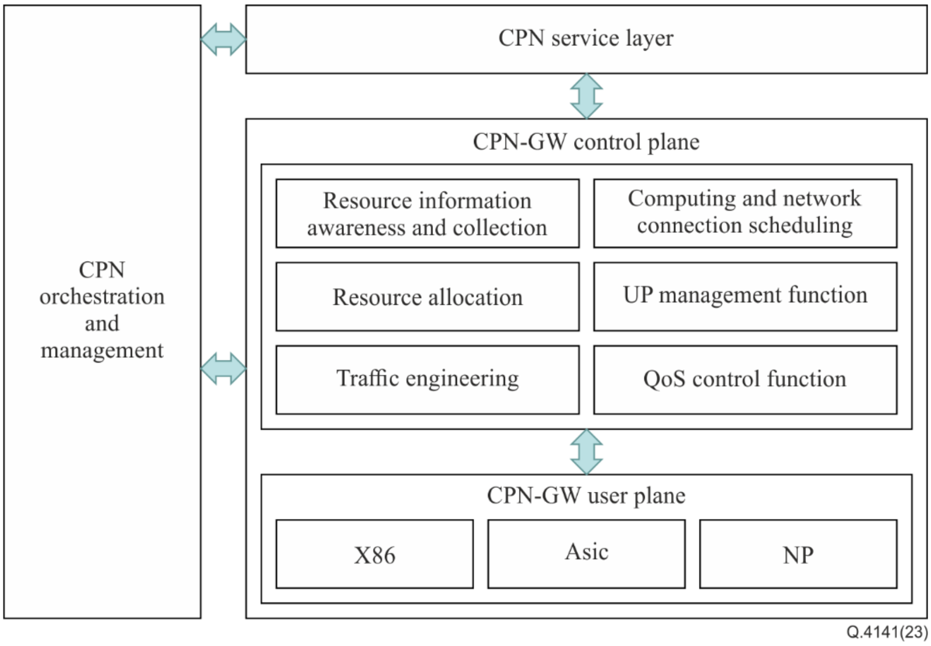

Figure 7-1 – Architecture of intelligence control for the CPN gateway

The CPN gateway should include the CPN gateway control plane and user plane, as well as interact with the CPN orchestration and management, and the components of the CPN service layer to realize the intelligent control for the CPN gateway. The intelligent control of the CPN gateway mainly realizes the routing and scheduling of computing power resources, while combining with the orchestration and management of the CPN gateway to achieve controllable, manageable, and schedulable computing power resources and services in the network. Specifically, the CPN orchestration and management mainly provides a unified management system for the CPN and computing power, including computing power measurement and modelling, computing power registration, and computing power operation administration and maintenance (OAM), etc. The CPN gateway control plane is mainly based on perception of network resource and computing power resource information for integrated scheduling of the CPN, also including functions such as computing power perception and notification, and computing power routing control, etc. The CPN gateway user plane is mainly based on the computing power routing table to transmit users' data. It further includes functions such as computing power application perception and computing power monitoring, etc.

7.1 CPN gateway control plane

The CPN gateway control plane introduces computing power information into the routing domain for computing power perceiving, routing control and scheduling, which mainly includes the computing power perception and the business scheduling functions. The computing power perception further includes computing power discovery and computing power notification functions, mainly used to collect computing power resource and network resource information. The business scheduling is mainly based on computing power and network status information to select appropriate service nodes to process user services.

Resource information awareness and collection is a function that gathers the computing resources and network resources in the connected computing power resource pool. The resource information table is maintained by this function. Moreover, the resource information table will be exchanged among the adjacent CPN gateways to assist the computing power scheduling. Also the resource information table will be sent in advance to the CPN service layer and CPN orchestration and management component for scheduling the policy of computing task offloading.

Resource allocation is a function that schedules the computing resources in the computing power resource pool. The function receives computing power transaction information from the CPN service layer and by looking up the resource information table, the resources scheduling policy will be made by this function. Additionally, when the user requests for the edge computing services with low latency requirements, the functions residing in the adjacent CPN gateways will cooperate to allocate the resources.

Computing and network connection scheduling is a function that builds up the connection between the computing power consumer and the computing power provider. The CPN service layer sends the network performance parameters such as bandwidth, latency, jitter, etc. to the function entity, and using the resource allocation policy, this function schedules the network connection to meet the various computing service demands.

7.2 CPN gateway user plane

The CPN gateway user plane mainly provides user packets transmission under the instruction of the CPN gateway control plane. The CPN gateway user plane can be implemented in different types of forwarding hardware, including network process (NP)-based dedicated equipment, application specific integrated circuit (ASIC)-based dedicated equipment, or X86-based commercial equipment, etc. Multiple routing and forwarding schemes have been proposed based on mainstream IPv6/SRv6 protocol extensions, including edge update, software-defined networking (SDN) control, and hop by hop update. The CPN gateway user plane should support multiple methods of traditional IP addressing and IP/service ID hybrid addressing, forming a new data plane for computing power routing.

The CPN gateway user plane should have the following the capabilities:

1) Support forwarding packets according to the computing power forwarding table items.
2) Support scheduling of available computing resources of computing resource pool and associated network resources.
3) Support reporting network status information such as node capabilities, neighbour relationships, and IP prefixes to the CPN gateway control plane, and executing instructions issued by the CPN gateway control plane and forwarding packets.

4) Support forwarding packets according to the different forwarding policies.

5) Support actively establishing network connections between designated users and computing power resource pools according to the service network connection demands issued by the CPN gateway control plane. The connection can be created using various technologies, such as virtual extensible local area network (VxLAN), SR/SRv6, secure socket layer (SSL) virtual private network (VPN), IPsec VPN, etc.

7.3 CPN orchestration and management

CPN orchestration and management supports the unified orchestrations of computing resources, network resources, and computing services by managing them. The orchestration function includes unified orchestration of computing power resources and network resources, service intention perception analysis, digital twins of computing networks, and intelligent computing power resource and network resource orchestration.

The main function of the CPN orchestration and management is to orchestrate computing and network resources, quickly allocate resources based on transaction contracts, such as virtual machine/container allocation, virtual private cloud (VPC) construction, network connection establishment, etc., and quickly recover resources and update resource information after completion. In addition, it supports computing power modelling according to various services and operation administration and maintenance (OAM) functions for the CPN [ITU-T Y.2501].

7.4 CPN service layer

The CPN service layer is responsible for the information exchange between the computing network and user services. User services, such as VR/AR, V2X, AI, etc. support submitting their own computing power and network access requests to the computing power network, as well as understanding real-time status information of the computing power resources and network resources, to achieve on-demand provision and flexible scheduling of computing power resources and network resources. Specifically, the service requests of the business or application are mapped to service application information and user business requests, including parameters such as computing power requests, and sent to the computing power routing node. The CPN service layer can also realize the functions of the computing power transaction, it supports resource information processing, billing and transaction process execution [ITU-T Q.4140].

8 Signalling procedures of intelligence control for the CPN gateway

8.1 Information interaction for intelligence control for the CPN gateway

In CPN, the CPN gateways exchange computing power resource information and network resource information with each other and with the forwarding equipment of the bearer network through control signalling, and then report them to the CPN gateway control component. The CPN gateway control component generates forwarding routes according to the computing resource information, network topology information and network status information, combined with forwarding strategies, and then selects appropriate computing nodes to deliver tasks. To this end, further discussion is needed on the relevant content of information exchanged, as well as the technical implementation requirements. The architecture is shown in Figure 8-1.

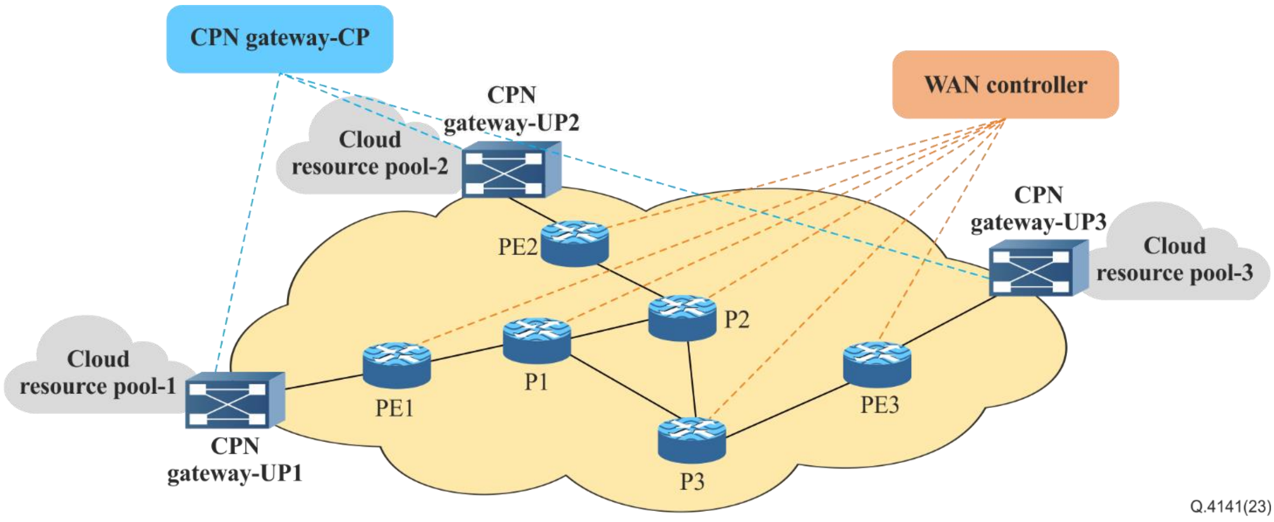

Figure 8-1 – Information exchange in the CPN

1) Contents of information interaction

The two-way transmission of information to each other is realized through information collection and information distribution. The interactive information includes network information and computing power resources information.

Network information:

– Topology information of the network.

– Link bandwidth, delay, jitter, packet loss rate and other related information.

Computing power resources information:

– CPU, graphics processing unit (GPU), neural network processing unit (NPU) and other processor information.

– Memory information.

– Store information.

2) Technical requirements for information interaction

- Information exchange between computing power network gateway user plane (CPN-GW-UP) and provider edge (PE) equipment:
  - The CPN gateway supports the proactive notification of the latest status of the internal information of the cloud resource pool to the bearer network PE equipment on a regular basis.
  - Provider edge (PE) equipment supports actively requesting the CPN gateway for internal information of the cloud resource pool.
  - The CPN gateway supports the ability to adjust the frequency of advertising information to PE equipment.
- Information exchange between bearer network PE equipment:
  - Support information automatic discovery and information exchange such as based on interior gateway protocol (IGP), border gateway protocol (BGP), and BGP+ Ethernet virtual private network (EVPN).
  - Supports the ability to adjust the frequency of information announcements.
- Information exchange between bearer network equipment and the computing power network gateway control plane (CPN-GW-CP)/PE controller:

– The bearer network equipment supports reporting information to the CPN-GW-CP/PE controller.

– The ability of the CPN-GW-CP/PE controller to support routing decisions and path policy issuance.

– The CPN-GW-CP/PE controller supports information exchange with bearer network equipment, and realizes the ability to make path decisions and deliver results.

8.2 Signalling procedures for intelligence control for the CPN gateway

8.2.1 Computing power resource information perception process

The perception of computing power resource information usually includes the IP address, computing power, storage capacity, and other contents of the computing power resource pool. The cloud resource pools are generally centrally managed by the cloud management platforms. The CPN gateway can interact with the cloud management platform through interfaces (such as RESTful APIs) to obtain the IP address, computing power, storage capacity, etc. of the computing resource pool, and ultimately report the perceived computing power information to the CPN-GW-CP. The procedures are shown in Figure 8-2.

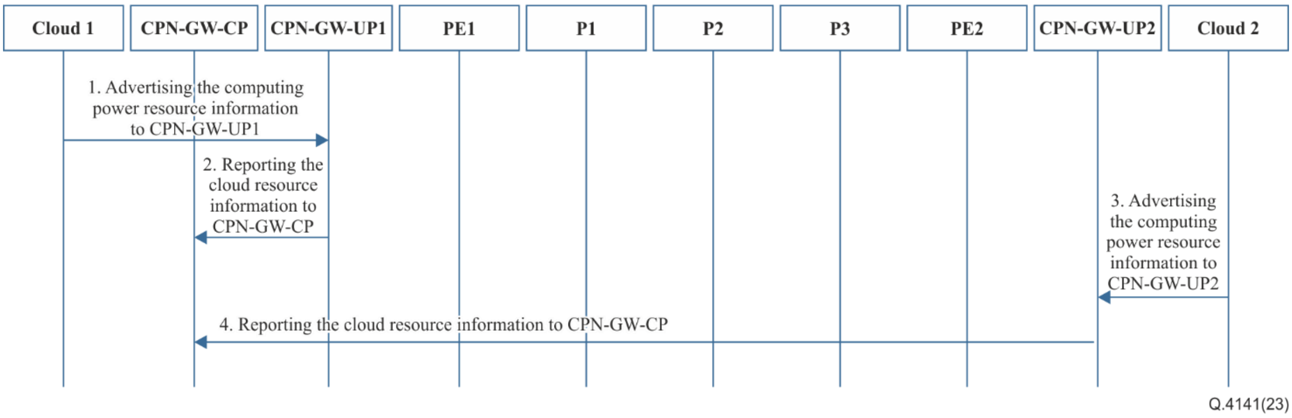

Figure 8-2 – Computing power information perception process

Step 1: Cloud resource pools 1 advertises the computing power resource information to CPN-GW-UP1.

Step 2: CPN-GW-UP1 reports the cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 3: Cloud resource pools 2 advertises the computing power resource information to CPN-GW-UP2.

Step 4: CPN-GW-UP2 reports cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

8.2.2 Network information perception process

The perception of network information usually includes content such as latency, bandwidth, packet loss rate, jitter, etc. Taking network latency as an example, due to the distribution of computing power resource pools in different locations, the network path from users to resource pools will also change based on network congestion status. Therefore, it is necessary to detect the latency information between users and various computing power resources pools. The delay detection of the CPN gateway is divided into two parts: one is the delay detection between the computing power resource pool and the CPN gateway. The other is the delay detection between the CPN gateways. The procedures are shown in Figure 8-3.

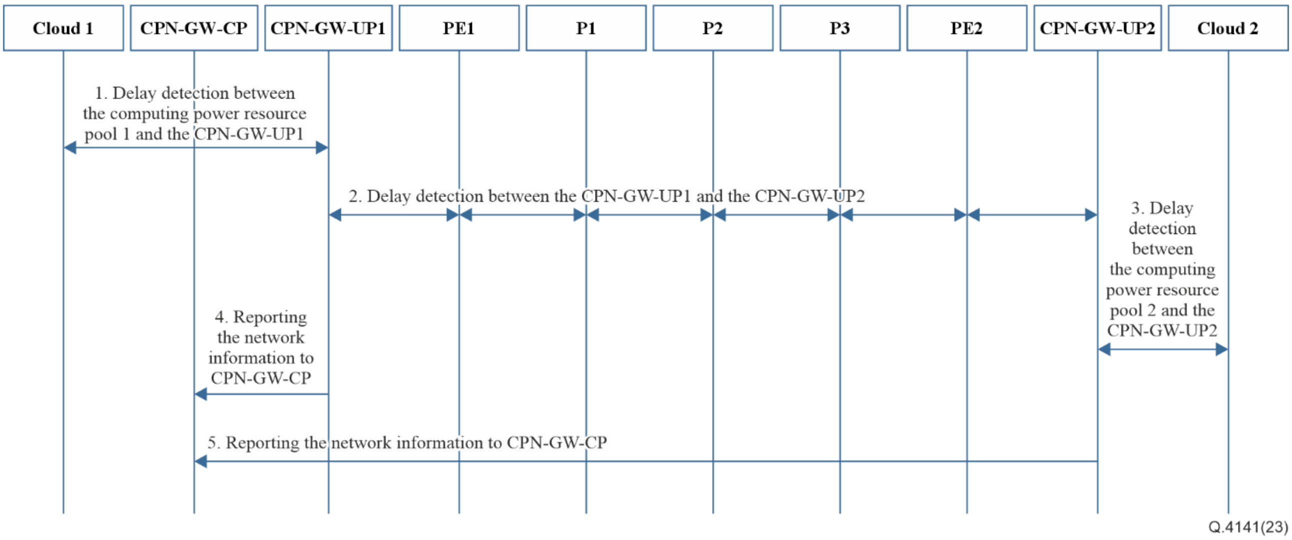

Figure 8-3 – Network information perception process

Step 1: Delay detection between the computing power resource pool 1 and the CPN gateway CPN-GW-UP1.

Step 2: Delay detection between the CPN gateway CPN-GW-UP1 and the CPN gateway CPN-GW-UP2.

Step 3: Delay detection between the computing power resource pool 2 and the CPN gateway CPN-GW-UP2.

Step 4: CPN-GW-UP1 reports the network information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 5: CPN-GW-UP2 reports the network information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

8.2.3 Computing power routing strategy

The CPN-GW-CP collects computing power resource and network resource information through the CPN-GW-UP, and determines the optimal computing resource node and network path, and then issues computing resource allocation instructions and routing strategies to the CPN-GW-UP separately. The CPN-GW-UP transmits user packets under the instructions of the CPN-GW-CP.

The procedures are shown in Figure 8-4.

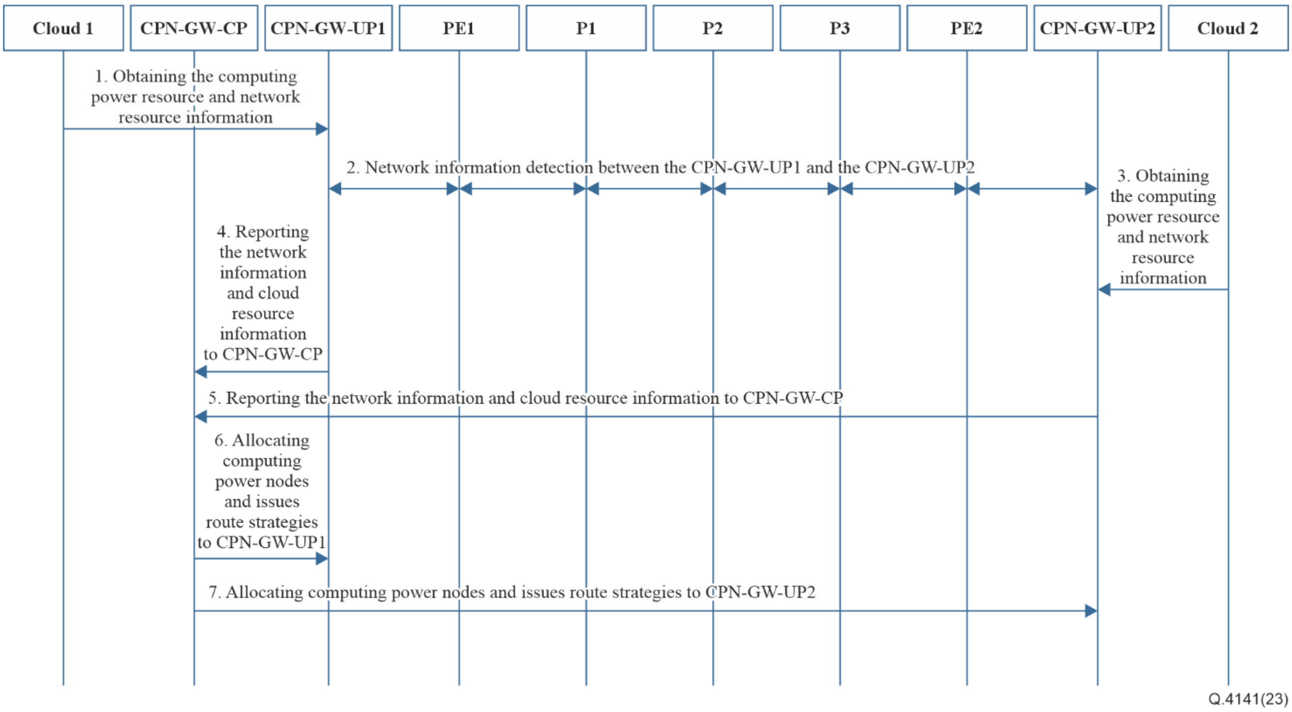

Figure 8-4 – Computing power routing strategy

Step 1: CPN-GW-UP1 obtains the computing power resource and network resource information.

Step 2: Network information detection between the CPN-GW-UP1 and CPN-GW-UP2.

Step 3: CPN-GW-UP2 obtains the computing power resource and network resource information.

Step 4: CPN-GW-UP1 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 5: CPN-GW-UP2 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 6: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP1 according to service requirements.

Step 7: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP2 according to service requirements.

8.2.4 Route selection procedure

The CPN gateway can support the path selection function based on IP route forwarding, multiple protocol label switch (MPLS) label forwarding and the segment routing (SR) source route, and can control the order in which traffic enters the processing unit according to service requirements. The procedures are shown in Figure 8-5.

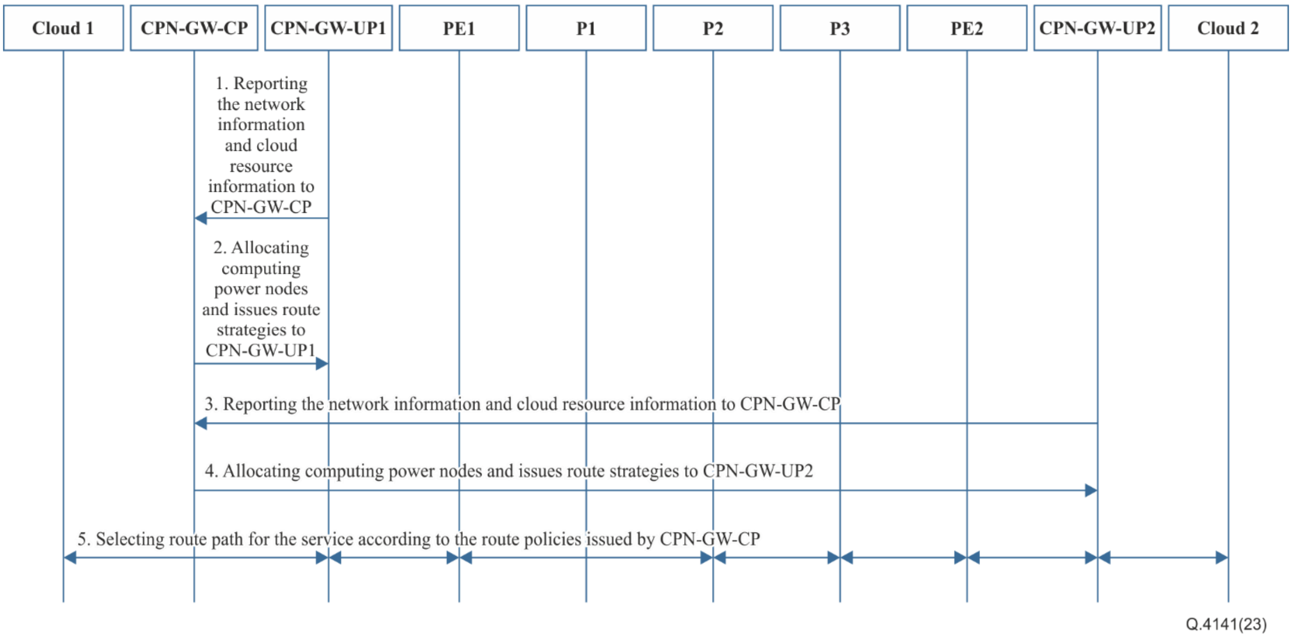

Figure 8-5 – Route selection procedure

Step 1: CPN-GW-UP1 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 2: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP1 according to the service requirements.

Step 3: CPN-GW-UP2 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 4: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP2 according to the service requirements.

Step 5: CPN-GW-UP1 selects cloud1 → CPN-GW-UP1 → PE1 → P2 → P 3→ PE2 → CPN-GW-UP2 → cloud2 path for forwarding packets according to the route policies issued by CPN-GW-CP.

8.2.5 Load balancing process

The CPN gateway can support multi-egress functions based on IP route forwarding, MPLS label forwarding and SR source route, and can balance traffic among multiple processing units according to the service requirements. The procedures are shown in Figure 8-6.

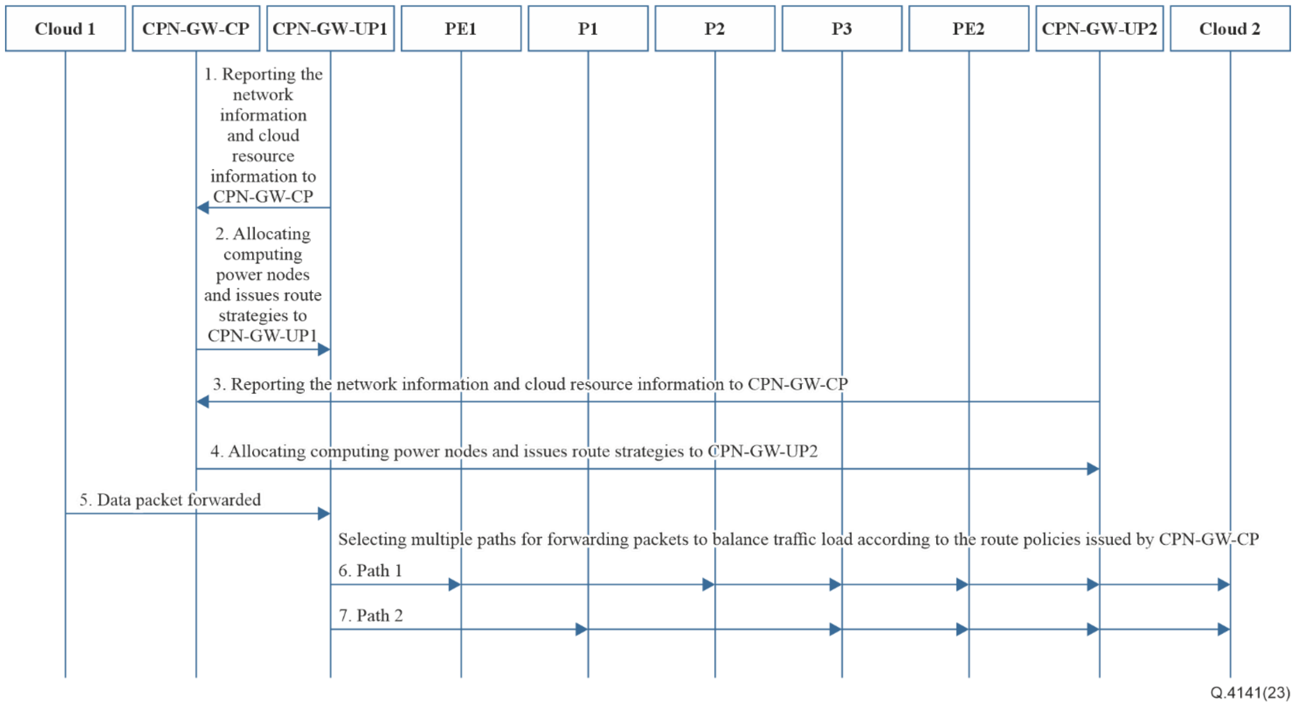

Figure 8-6 – Load balancing procedure

Step 1: CPN-GW-UP1 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 2: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP1 according to the service requirements.

Step 3: CPN-GW-UP2 reports the network information and cloud resource information to CPN-GW-CP, CPN-GW-CP maintains the resource information table.

Step 4: CPN-GW-CP allocates computing power nodes and issues route strategies to CPN-GW-UP2 according to the service requirements.

Step 5-7: When the data packet is forwarded from cloud1 to CPN-GW-UP1, CPN-GW-UP1 selects multiple paths for forwarding packets to balance traffic loads according to the route policies issued by CPN-GW-CP. For example:

– Path 1: CPN-GW-UP1 → PE1 → P2 → P3 → PE2 → CPN-GW-UP2 → cloud2, and

– Path 2: CPN-GW-UP1 → PE1 → P3 → PE3 → PE2 → CPN-GW-UP2 → cloud2.

9 Security considerations

The new architecture, technologies, and services of computing power networks may pose new security risks that need to be addressed, and new security mechanisms that are compatible with CPNs are needed to ensure their security. There are potential complex network risks and computing power node security risks in the infrastructure layer. The orchestration and management layer are involved in the security risks of orchestration and the issue of uncontrolled use of computing power. The CPN service layer faces issues such as accessing malicious nodes, untrustworthy transactions, and insecure applications. In addition, there may be data security risks in the computing power network, such as unreliable calculation results and uncontrollable data flow. It is necessary to strengthen security through an integrated full process trust mechanism.

The computing network architecture puts forward new requirements for security. Intelligence control for the CPN gateway needs to build end-to-end multi network elements and multi-level cooperative security protection capabilities. Information security, functional security, privacy security, trusted security and demand characteristics need to be considered as a whole to provide different security characteristics at different levels. The specific requirements are as follows:

– The security function shall support intelligent and active adaptation architecture of intelligence control for the CPN gateway, and be able to deploy and expand flexibly.
– Security functions and service support are dynamically deployed to the CPN gateway with applications.
– The CPN gateway supports the isolation of the network domain, controls the scope of security attacks and risks, and avoids attacks from point to surface.
– The CPN gateway supports continuous security detection and response.

Appendix I

Use case descriptions

(This appendix does not form an integral part of this Recommendation.)

I.1 Edge computing interconnect scenario

A use case of an edge computing interconnect scenario is shown in Figure I.1. The consumers could be Internet of things (IoT) devices, mobile device users, auto-driving vehicles, etc. The computing task requiring low communication latency will be distributed to the nearest edge cloud by the CPN transaction platform in advance. The platform will also allocate other edge clouds to the computing task for the candidate computing node. When the edge cloud lacks computing power, the CPN gateways connected to the computing node should schedule inter-domain computing power in a distributed way. As shown in Figure I.1, the edge cloud1 is short of computing power. The scheduling procedure is described below.

Procedure 0 – Information on the computing resources and network performances in the edge cloud is reported to the control plane of the CPN gateway and an intra-domain resources information table is maintained by the resource information collection function. The table will be exchanged by the neighbouring CPN-GW-CP1 and CPN-GW-CP2 to maintain the inter-domain resource information table.

Procedures 1 and 2 – The consumers submit a computing task to the edge cloud1. However, the computing power is not enough. An inter-domain computing power scheduling request will be sent to the connected CPN-GW-CP1.

Procedure 3 – According to the inter-domain resource information table, the edge cloud1 chooses edge cloud2 to conduct computing offloading and sends the computing task.

Procedure 4 – The CPN-GW-CP2 consults the intra-domain resource information table and allocates the computing power for the task. The network connection will then be constructed to satisfy the network requirements of the task.

Procedure 5 – The consumers will be notified that the computing node is changed and send data to the servers deployed in the edge cloud2. The results will then be processed and returned to the consumers.

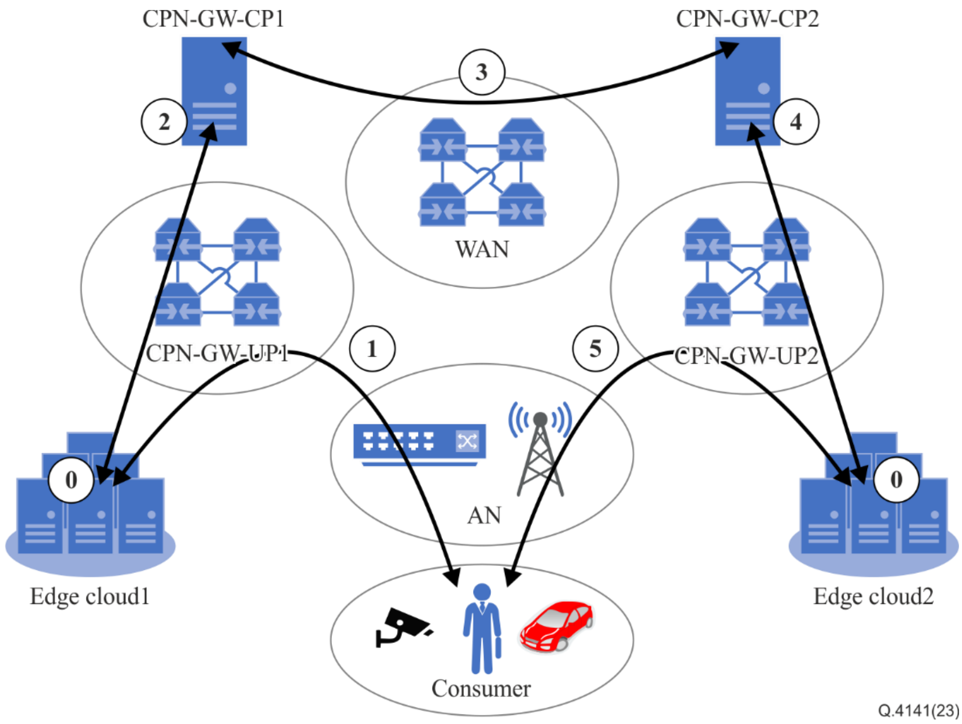

Figure I.1 – Use case in an edge computing interconnect scenario

I.2 Data centre interconnect scenario

As shown in Figure I.2, a use case in a data centre interconnect scenario is described. Unlike in the edge computing interconnect scenario, the network control plane is divided into two parts. The inter-domain part is the wide area network (WAN) inter-domain controller, which manages the inter-domain traffic. The intra-domain part is the intra-domain control plane of the CPN gateways, which manages the intra-domain traffic. Thus, in this scenario, the inter-domain computing power is scheduled in a centralized way and the procedure is described below.

Procedure 0 – The intra-domain information of computing resources and network performances will be collected by the intra-domain control plane of the CPN gateway and the intra-domain resource information table is maintained in the CPN gateway control plane.

Procedure 1 – The intra-domain resource information table is sent to the WAN inter-domain controller. The controller maintains the inter-domain resource information table, which contains the inter-domain information of computing resources and network performances.

Procedure 2 – The inter-domain resource information table is sent to the CPN transaction platform and the platform integrates the resource information and makes a resource quotation.

Procedures 3 and 4 – When a user terminal requests resources for a computing task, the request will be sent to the platform. The platform then generates a resource view for the user. The users choose the most suitable resources and make a transaction contract with providers.

Procedure 5 – The platform sends the transaction information to the WAN inter-domain controller. The controller consults the inter-domain resources information table and allocates the inter-domain resources for the transaction. The inter-domain network connection between the user and the chosen data centre will be constructed and maintained by the inter-domain controller. The transaction will then be delivered to the intra-domain CPN gateway controller.

Procedure 6 – The transaction will be delivered to the intra-domain CPN gateway controller.

Procedure 7 – By consulting the intra-domain resource information table, resources will be allocated locally by the intra-domain CPN gateway controller and the network connection between the resources and the consumers will be constructed and maintained.

Procedures 8 and 9 – The user terminal sends the computing task to the server. The server accomplishes the computing task and returns the results to the user terminal.

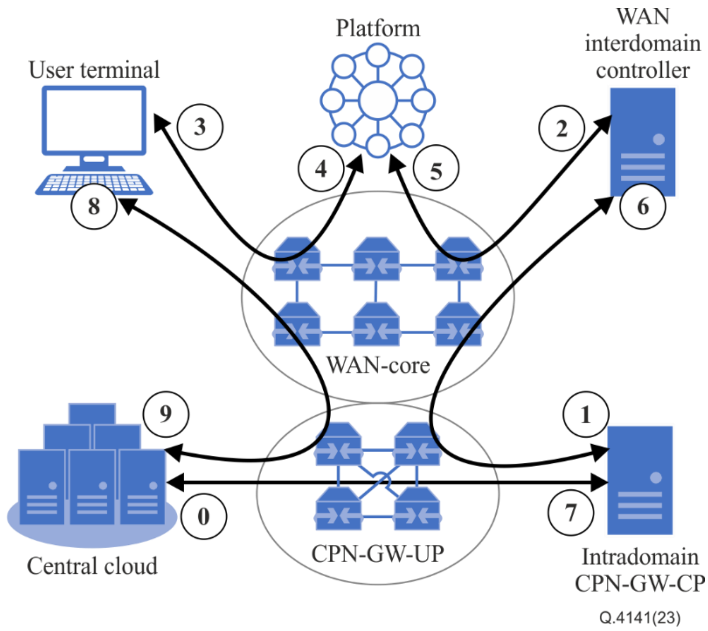

Figure I.2 – Use case in a data centre interconnect scenario

Appendix II

Deployment of the CPN gateway

(This appendix does not form an integral part of this Recommendation.)

In the application deployment of CPN gateways, two networking solutions, hybrid and distributed, are mainly considered.

II.1 CPN gateway hybrid networking

In the hybrid networking solution, the computing network orchestration system [ITU-T Y.2501] relies on the cloud/computing management and control module to collect computing resource information from each computing power resource pool through the CPN gateway, and collects network topology information through the network management and control module. The computing network orchestration system determines the optimal computing resource nodes and network forwarding paths. The cloud/computing management and control module, and the network management and control module issue computing resource allocation instructions and routing strategies respectively.

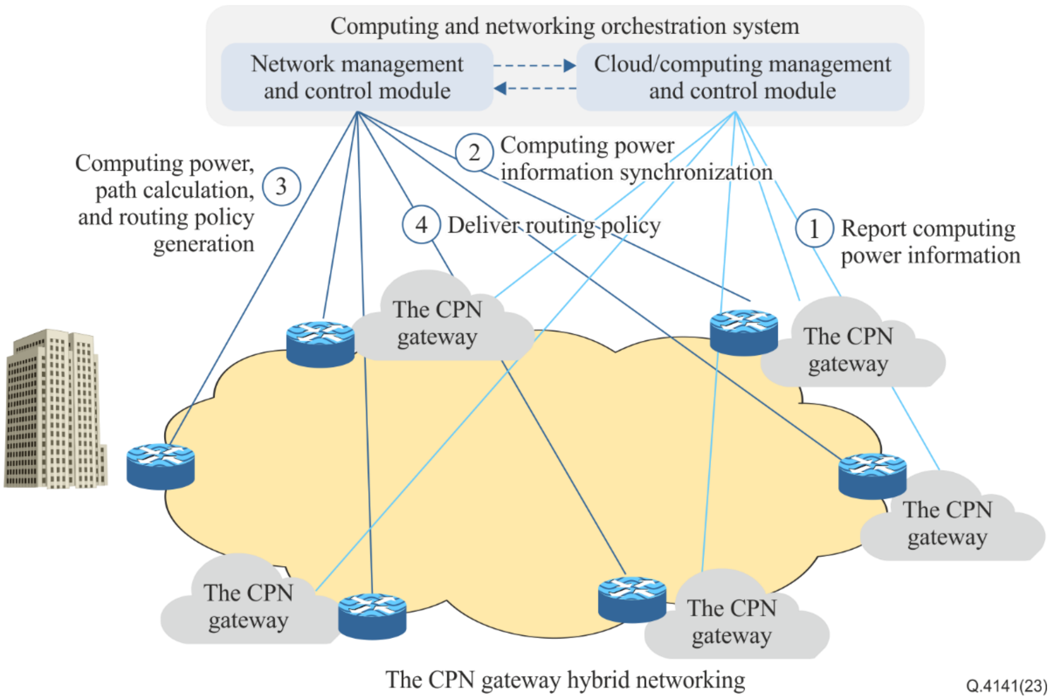

Figure II.1 – The CPN gateway hybrid networking

Under this architecture, the main functions of the CPN gateway should include obtaining computing resource information and link information of computing power nodes, and receiving path policy information issued by the network management and control module, etc.

II.2 CPN gateway distributed networking

In the distributed solution, the CPN gateway needs to implement functions such as computing power resource awareness, computing power resource information distribution, computing resource routing table generation, and policy customization, as shown in Figure II.2. In addition to supporting the release and notification of computing resource information, the distributed solution also needs to generate routing policies through computing power and path calculations and bind the routing policies based on user and application perception, thereby achieving information synchronization of computing resources and network resources, and unified scheduling.

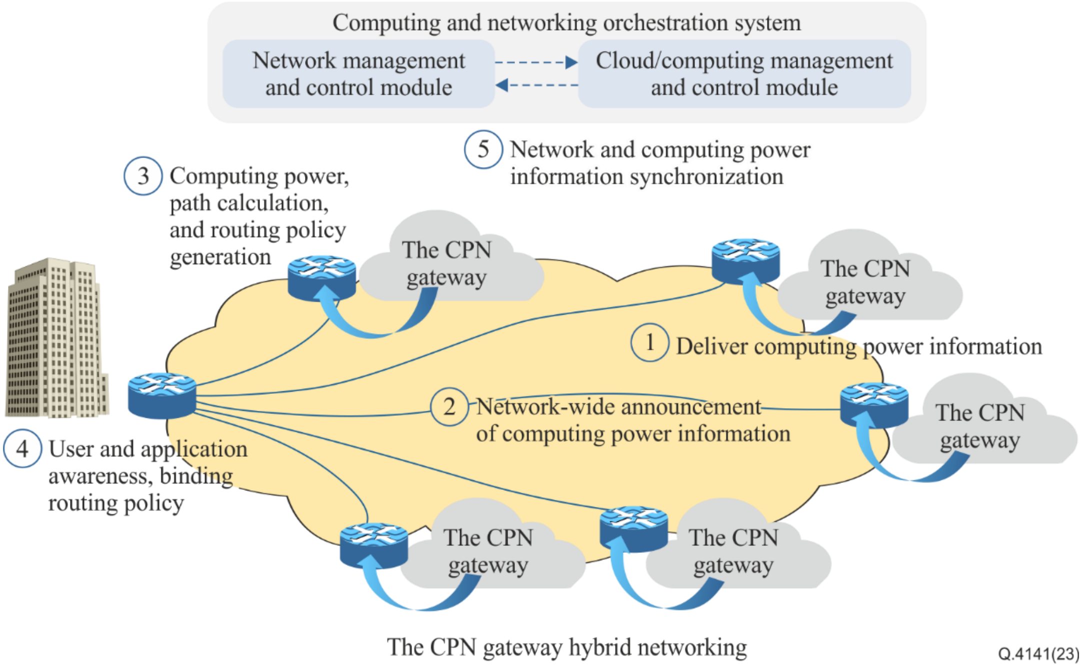

Figure II.2 – CPN gateway distributed networking

The CPN gateway can sense the computing power resource pool information, and sense network information such as delay and packet loss through telemetry and other technologies, and so form a network delay circle for users to choose the appropriate computing power resource pool. The information is transmitted, using for example border gateway protocol (BGP) or internal gateway protocol (IGP) extension, to each computing power routing node device and computing power gateway in the computing power domain, to form a computing power IP routing table and build a resource view.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	**Switching and signalling, and associated measurements and tests**
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2024