ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T Y.3540 (10/2023)

SERIES Y: Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

Cloud Computing

Edge computing – Overview and high-level requirements

ITU-T Y-SERIES RECOMMENDATIONS

Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

GLOBAL INFORMATION INFRASTRUCTURE	Y.100-Y.999
General	Y.100-Y.199
Services, applications and middleware	Y.200-Y.299
Network aspects	Y.300-Y.399
Interfaces and protocols	Y.400-Y.499
Numbering, addressing and naming	Y.500-Y.599
Operation, administration and maintenance	Y.600-Y.699
Security	Y.700-Y.799
Performances	Y.800-Y.899
INTERNET PROTOCOL ASPECTS	Y.1000-Y.1999
General	Y.1000-Y.1099
Services and applications	Y.1100-Y.1199
Architecture, access, network capabilities and resource management	Y.1200-Y.1299
Transport	Y.1300-Y.1399
Interworking	Y.1400-Y.1499
Quality of service and network performance	Y.1500-Y.1599
Signalling	Y.1600-Y.1699
Operation, administration and maintenance	Y.1700-Y.1799
Charging	Y.1800-Y.1899
IPTV over NGN	Y.1900-Y.1999
NEXT GENERATION NETWORKS	Y.2000-Y.2999
Frameworks and functional architecture models	Y.2000-Y.2099
Quality of Service and performance	Y.2100-Y.2199
Service aspects: Service capabilities and service architecture	Y.2200-Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250-Y.2299
Enhancements to NGN	Y.2300-Y.2399
Network management	Y.2400-Y.2499
Computing power networks	Y.2500-Y.2599
Packet-based Networks	Y.2600-Y.2699
Security	Y.2700-Y.2799
Generalized mobility	Y.2800-Y.2899
Carrier grade open environment	Y.2900-Y.2999
FUTURE NETWORKS	Y.3000-Y.3499
CLOUD COMPUTING	**Y.3500-Y.3599**
BIG DATA	Y.3600-Y.3799
QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800-Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	Y.4000-Y.4999
General	Y.4000-Y.4049
Definitions and terminologies	Y.4050-Y.4099
Requirements and use cases	Y.4100-Y.4249
Infrastructure, connectivity and networks	Y.4250-Y.4399
Frameworks, architectures and protocols	Y.4400-Y.4549
Services, applications, computation and data processing	Y.4550-Y.4699
Management, control and performance	Y.4700-Y.4799
Identification and security	Y.4800-Y.4899
Evaluation and assessment	Y.4900-Y.4999

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

Recommendation ITU-T Y.3540

Edge computing – Overview and high-level requirements

Summary

Recommendation ITU-T Y.3540 gives an overview and high-level requirements for edge computing. For edge computing, Recommendation ITU-T Y.3540 defines related terms, and gives an overview of its concept, common characteristics and operational ecosystem with main roles. Recommendation ITU-T Y.3540 also describes orchestration aspects for edge computing and its relationship with other technologies. Recommendation ITU-T Y.3540 provides high-level requirements through various use cases.

History*

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T Y.3540	2023-10-22	13	11.1002/1000/15538

Keywords

Edge computing, edge computing ecosystem, high-level requirements.

* 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 Y.3540

Edge computing – Overview and high-level requirements

1 Scope

This Recommendation gives an overview and high-level requirements for edge computing. It addresses the following:

– an overview of edge computing;
– an overview of the roles of individual actors in edge-computing operation;
– a technical view of edge computing;
– high-level requirements for edge computing.

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.3073] Recommendation ITU-T Y.3073 (2019), *Framework for service function chaining in information-centric networking*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 cloud computing [b-ITU-T Y.3500]: Paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand.

NOTE – Examples of resources include servers, operating systems, networks, software, applications and storage equipment.

3.1.2 cloud service [b-ITU-T Y.3500]: One or more capabilities offered via cloud computing invoked using a defined interface.

3.1.3 cloud service provider [b-ITU-T Y.3500]: Party which makes cloud services available.

3.1.4 edge computing [ITU-T Y.3073]: This refers to a strategy to deploy processing capability at network edge where end terminals are connected, and to perform the processing of data which is derived from and fed to the end terminals.

3.1.5 Internet of things (IoT) [b-ITU-T Y.2060]: A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.

NOTE 1 – Through the exploitation of identification, data capture, processing and communication capabilities, the IoT makes full use of things to offer services to all kinds of applications, whilst ensuring that security and privacy requirements are fulfilled.

NOTE 2 – From a broader perspective, the IoT can be perceived as a vision with technological and societal implications.

3.1.6 party [b-ISO 27729]: Natural person or legal person, whether or not incorporated, or a group of either.

3.1.7 role [b-ITU-T Y.3502]: A set of activities that serves a common purpose.

3.1.8 quality of service [b-ITU-T Q.1743]: The collective effect of service performances, which determine the degree of satisfaction of a user of a service. It is characterized by the combined aspects of performance factors applicable to all services, such as:

– service operability performance;

– service accessibility performance;

– service retainability performance;

– service integrity performance; and

– other factors specific to each service.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 edge computing application: Business logic deployed and delivered to users using edge computing service.

3.2.2 edge computing customer: A party corresponding to a natural person or an entity in a business relationship to use an edge computing application or edge computing service.

3.2.3 edge computing infrastructure: Resources that provide processing, network and storage at or near a physical location close to user in edge computing.

3.2.4 edge computing partner: A party that supports or assists an edge computing provider or an edge computing customer, or both.

3.2.5 edge computing provider: A party providing edge computing application available to edge computing customer, and edge computing services available to edge computing customer and edge computing partner.

3.2.6 edge computing service: One or more capabilities provided on edge computing infrastructure.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

AR Augmented Reality

CCTV Closed Circuit Television

CPN Computing Power Network

ECC Edge-Computing Customer

ECN Edge-Computing Partner

ECP Edge-Computing Provider

HD High Definition

IMT-2020 International Mobile Telecommunication 2020

IoT Internet of Things

IP Internet Protocol

MEC	Multi-access Edge Computing
NVMe	Non-Volatile Memory express
PaaS	Platform as a Service
QoS	Quality of Service
RAM	Random Access Memory
URL	Uniform Resource Locator
V2X	Vehicle to everything
VM	Virtual Machine
VR	Virtual Reality
WiFi	Wireless Fidelity

5 Convention

The phrase "is required" indicates a requirement that must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

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

In the body of this Recommendation and its annexes, the words "shall", "shall not", "should" and "may" sometimes appear, in which case they are to be interpreted, respectively, as "is required", "is prohibited from", "is recommended", and "can optionally". The appearance of such phrases or keywords in an appendix or in material explicitly marked as informative are to be interpreted as having no normative intent.

6 Overview of edge computing

6.1 Concept of edge computing

Edge computing is a computing technology to deploy processing capability at a network edge to which end terminals are connected, and to process data that is derived from and fed to the end terminals [ITU-T Y.3073]. A network edge corresponds to a physical location where data are generated or consumed.

Figure 6-1 shows edge computing in terms of the physical location of the computing process. Edge computing users include individuals, applications or systems that generate data (e.g., sensors at the edge of networks). Edge computing users differ from the traditional user of a network because they use applications and services that are deployed as close to them as possible.

Edge computing comprises an application, service and infrastructure, which are described in clauses 6.1.1 to 6.1.3.

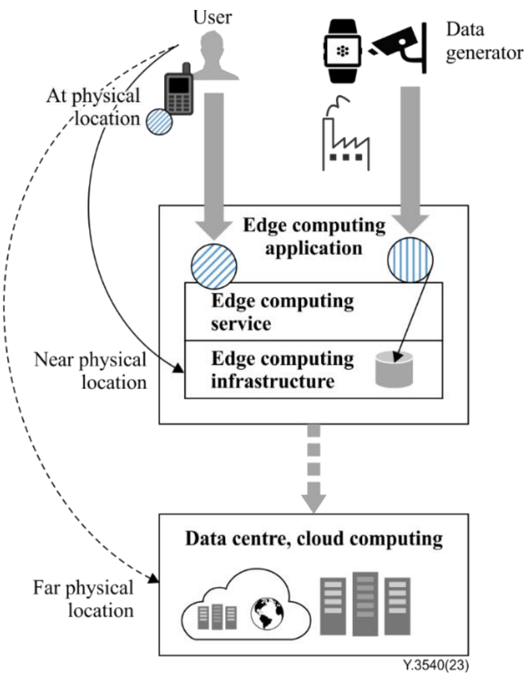


Figure 6-1 – A conceptual diagram of edge computing

6.1.1 Edge-computing application

In edge computing, an application is business logic delivered to users using a service at or near a physical location. Furthermore, the application is then deployed on infrastructure using services in edge computing.

An edge-computing application is used in various application domains such as smart factory, gaming, autonomous vehicle, transportation, smart city, smart retail, smart robot, smart home, healthcare, smart agriculture, smart grid and smart building (see Appendix I).

Examples of edge-computing applications are described as follows.

- Autonomous vehicle: edge computing enables autonomous vehicles: to make real-time decision without the help of a data centre; improves safety; and reduces latency.
- Smart city: edge computing is used to monitor and manage smart city infrastructure such as traffic signal control, smart parking and energy consumption.
- Retail: edge computing enables retail to analyse customer data in real time, allowing target marketing and improvement in the customer shopping experience.
- Healthcare: edge computing processes patient data for personal care, improving the speed and accuracy of diagnosis, treatment and patient monitoring.
- Manufacturing: edge computing improves manufacturing processes by enabling real-time monitoring and optimization of production line.

An edge-computing application processes user proximity data in real time and low latency over the network. For low latency, an edge-computing application uses a high-speed network to connect to the user and maintain seamless connection between that application and the user end terminal.

6.1.2 Edge-computing service

In edge computing, a service is one or more capabilities to provide an application on infrastructure, as shown in Figure 6-2.

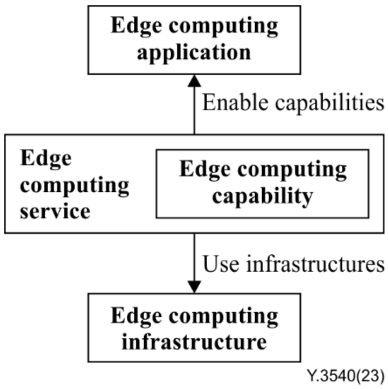


Figure 6-2 – The concept of edge-computing service

In edge computing, a service provides network and computing capabilities based on requirements and requests from applications.

The typical capabilities of an edge-computing service include the following.

– **User proximity capability**: deployment of an edge-computing application at or near physical location to reduce latency.

– **Network connectivity capability**: connection of pieces of edge-computing and other infrastructure to provide a service and application in edge computing.

– **Data proximity capability**: storage and processing of data at or near edge-computing applications.

– **Cooperation capability**: sharing resources in infrastructure, load balancing and offloading of an application in edge computing.

NOTE – As an example, the offloading of an edge-computing application includes transferring the role of resource-intensive computational processing to a data centre or cloud computing with a hardware accelerator or larger-scaled computing resources.

– **Orchestration capability**: in edge computing, coordination and management of a service and infrastructure as well as deployment, execution, control, and scheduling an application.

In edge computing, a service is possibly combined with other services to provide capabilities. An edge-computing service uses services from other infrastructure such as cloud service.

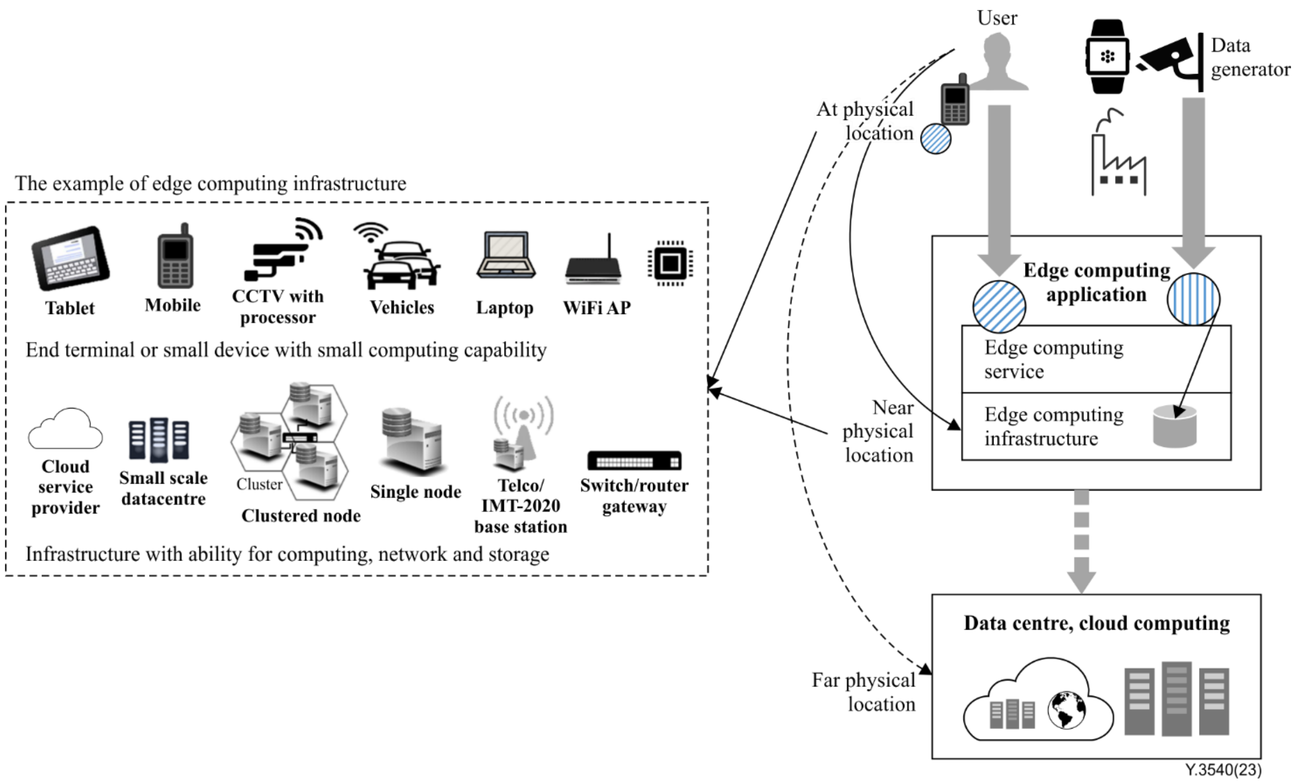
6.1.3 Edge-computing infrastructure

Edge-computing infrastructure is the physical or software resources at or near a physical location where edge-computing applications run and store data. Examples of edge-computing infrastructure are shown in Figure 6-3.

Edge-computing infrastructure includes a set of physical and software resources, the connected form of resources and other technology infrastructure (e.g., cloud computing and data centre).

– Physical resources include a single node, laptop, embedded system device or mobile device and small-scale data centre.

– Software resources include software programs, operating system, software platform, virtual resources such as container and virtual machine or virtual resource provided by a cloud service provider.



IMT-2020: International Mobile Telecommunication 2020; CCTV: closed circuit television; WiFi: wireless fidelity

Figure 6-3 – An example of edge-computing infrastructure

Edge-computing infrastructure has resources to perform edge-computing services. For example, in edge computing, user proximity capability uses infrastructure to deploy an application at or near a physical location close to user. Furthermore, network connectivity capability uses a network for connection among pieces of edge-computing infrastructure. Furthermore, data proximity capability uses data storage at the network edge. Cooperation capability uses high-performance resources for storage sharing, load balance and offloading of edge-computing applications.

NOTE 1 – High-performance resources include a graphics processing unit, field programmable gateway array, storage such as main memory, random access memory (RAM) and non-volatile memory express (NVMe).

NOTE 2 – According to clause 6 of [b-ITU-T Y.2501], the computing power network (CPN) is described as "a new type of network that realizes optimized resource allocation, by distributing computing, storage, network and other resource information of service nodes through a network control plane (such as a centralized controller, distributed routing protocol, etc.)". From the perspective of edge computing, the CPN is considered as a resource to provide network connectivity capability in infrastructure.

NOTE 3 – In edge computing, the resources available at the network edge are capable of steering traffic between users and applications. These network resources include switches, gateways, routers, security middleboxes and access points, and may be implemented in either hardware or software.

6.2 Common characteristics of edge computing

Common characteristics of edge computing are derived from its concept. The common characteristics of edge computing are as follows.

– **Low latency**: edge computing reduces the data transfer time between source and destination for its applications when response time is critical.

– **Network connection between infrastructure**: edge computing relies on reliable and fast network connectivity between its own and other infrastructure.

– **Data processing close to user**: edge computing processes data at or near the physical location where they are generated and consumed.

– **Minimal bandwidth consumption**: edge computing transfers minimal data to reduce network bandwidth. Edge computing shares and synchronizes data for minimal bandwidth consumption.
– **Scalability of infrastructure**: edge computing scales its infrastructure up or down when connecting to a large data centre or cloud computing.

NOTE 1 – According to various application domains, different characteristics of edge computing are identified. The justification of common characteristics of edge computing is described in Appendix I.

NOTE 2 – The common characteristics of edge computing provide improvement of its quality of service (QoS). Reducing latency, enabling nearby data processing and optimization of bandwidth usage improve QoS in terms of speed of use experience. Minimizing dependency on a centralized infrastructure and ensuring continuous operations on edge-computing infrastructure enhance QoS in reliability.

7 Ecosystem for edge computing

The edge-computing ecosystem describes the fundamental parties required to perform edge computing and the roles of their actions.

The main edge computing roles are as follows:

– edge-computing customer (ECC);
– edge-computing partner (ECN);
– edge-computing provider (ECP).

7.1 Edge-computing customer

In edge computing, an ECC is a party corresponding to a natural person or an entity in a business relationship to use an application or service. An ECC includes any user, application or system such as a generator to generate or consume data. An ECC includes an administrator for all operational processes of an edge-computing service to ensure use by its customers. An ECC includes a manager that is responsible for the use of an edge-computing application to meet a business goal. An ECC monitors services and infrastructure in edge computing.

7.2 Edge-computing partner

An ECN is a party that supports or assists an ECP or an ECC, or both. In edge computing, the ECN develops a service and application to enable its capability and integrate different services.

An ECN designs and creates capability of edge computing to develop its service. An ECN also enhances an edge-computing service. An ECN designs an interface for an ECC to use an application, service and infrastructure in edge computing.

7.3 Edge-computing provider

In edge computing, an ECP is a party making an application available to an ECC, and services available to an ECC and ECN. An ECP delivers, manages and operates, in edge computing, services and applications to ensure the provision of a service to an ECC via its infrastructure.

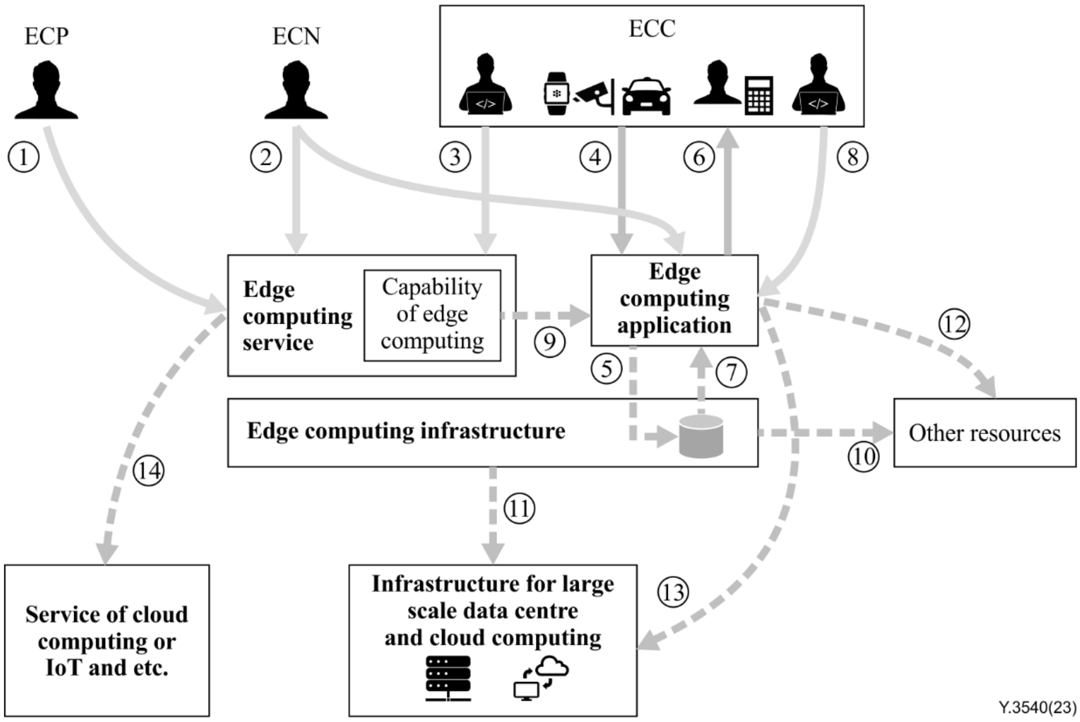
In edge computing, an ECP provides an application to meet the business goals of an ECC, and a service to operate, manage and develop the application using its service.

In edge computing, an ECP provides infrastructure, orchestrates a service and application, offers a service to an ECC, manages security, risks and service continuity for applications, handles customer requests, etc.

NOTE – In edge computing, [b-ITU-T Y.3123] provides a platform, which is a functional component to provide capabilities if three kinds, i.e., edge network, edge application and general information technology. From the perspectives of an ECP, an edge-computing platform is considered as one of its functional components.

7.4 Operations by main roles in edge computing

In edge computing, Figure 7-1 shows an example of operations on infrastructure, service and application by main roles.



①	In edge computing, an ECP provides a service with its capability
②	In edge computing, an ECN develops a service and application
③	In edge computing, an ECC) administers, controls and manages services provided to developers
④	In edge computing, an ECC uses an application to store and send data
⑤	In edge computing, an application stores data in infrastructure near their physical location
⑥	In edge computing, an ECC uses an application with minimization of data transfer
⑦	An edge-computing application processes or sends data on request from an ECC
⑧	An ECC controls and manages an edge-computing application
⑨	In edge computing, a service enables an application with its capabilities
⑩	Edge-computing infrastructure uses other resources to scale up and expand resources for network connectivity capability, if necessary
⑪	Edge-computing infrastructure uses other infrastructure for network connectivity capability, if necessary
⑫	An edge-computing application replicates or migrates to other resources for cooperation capability, if necessary
⑬	An edge-computing application offloads to a large-cale data centre for cooperation capability, if necessary
⑭	In edge computing, a service uses another service to support an application, if necessary

Figure 7-1 – Example of operations by main roles in edge computing

8 The technical view of edge computing

8.1 Orchestration aspects of edge computing

In edge computing, orchestration is the coordination and management of infrastructure, and ensures that infrastructure, services and applications work together seamlessly and efficiently to deliver according to user requirements. Examples of aspects of orchestration in edge computing include the following.

- **Infrastructure management**: management of pieces of edge-computing infrastructure, ensuring that they are properly configured and maintained, and managing their software updates and patches.

NOTE 1 – In edge computing, orchestration includes management of multiple pieces of infrastructure connected by software (e.g., agents or gateway software) or the management of a single resource within one piece.

- **Service management**: in edge computing, provision of services, management of service level agreements and monitoring services, e.g., for downtime and performance.

  NOTE 2 – Service management includes the discovery of an edge-computing application using its interface (Internet protocol (IP) address, domain name or uniform resource locator (URL)) for decentralized processing.

- **Data management**: management of large amounts of data that are generated at the edge, using data filtering, processing, aggregation and analysis to reduce the use of network resources.
- **Network management**: management of network connections between edge-computing and other infrastructure, including management of bandwidth, and ensuring network reliability and security.
- **Resource management**: monitoring and optimization of the use of computing, storage and network resources to ensure that an edge-computing service and application operate efficiently.
- **Security and privacy**: management of authentication, encryption and access controls to protect edge-computing application, service and infrastructure.
- **Application management**: management of the development, deployment and maintenance of an edge-computing application, ensuring that it operates reliably and efficiently.
- **Cooperation management**: management of task offloading, load balance and resource sharing of an edge-computing application to enhance its performance, reliability, and efficiency.

  NOTE 3 – Task offloading includes migration such as control process with checking edge-computing application status, checking the destination of edge-computing infrastructure, saving the current status of edge-computing application as a snapshot of application's image, and restoring the snapshot.

- **High-availability management**: for edge computing, provision of uninterrupted and reliable access to its application and service, even if hardware or software fails, a network is disrupted or other unexpected events occur.

  NOTE 4 – High-availability management provides a load balancer or high-availability proxy that periodically checks the connection, and performs a failover with a new connection using the service exposure if disconnected.

8.2 Relationship with other technologies

In edge computing, services of other technologies are used to provide its own service. Figure 8-1 illustrates the relationship with edge computing in terms of utilizing the services of other technologies such as cloud computing. As an example, edge computing utilizes cloud service (such as infrastructure as a service, platform as a service (PaaS) and software as a service) as its own service. Also, in IoT, services for an application, platform and network as described in [b-ITU-T Y.4100] are candidates as those that edge-computing services utilize for edge-computing capability.

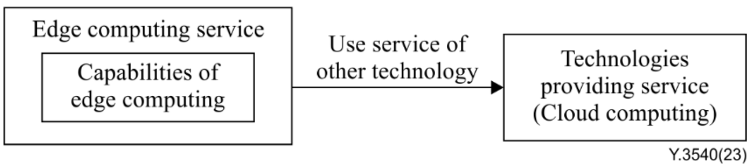


Figure 8-1 – The relationship in terms of the utilization of services

9 High-level requirements of edge computing

According to its application, the following high-level requirements of edge computing are considered. Not all applications of edge computing are needed to meet all its high-level requirements.

– **(User proximity)** It is required that edge computing provide its infrastructure at or near the physical location of an ECC to operate its application.

– **(Minimization of latency)** It is required that edge computing minimize latency with consideration of user proximity.

– **(Network connectivity for infrastructure)** It is required that edge computing provide network connection between pieces of edge-computing and other infrastructure such as cloud computing and data centres.

– **(Data affinity)** It is required that edge computing provide data close to its application.

– **(Mobility supports)** It is required that edge computing support its application to move or relocate within a network while maintaining its connectivity.

– **(Infrastructure interface)** It is required that edge computing provide interfaces with its infrastructure.

– **(Cooperation among applications)** It is required that edge computing cooperate with its applications by resource sharing, load balancing and task offloading.

– **(Resource provisioning)** It is required that edge computing provide edge-computing infrastructure for physical or software resources.

– **(User authentication)** It is required that edge computing authenticate users for access to its applications and services.

– **(Data privacy)** It is required that edge computing protect user data.

– **(Provision of service)** It is required that edge computing provide its service enabled with its capability.

– **(Orchestration**) It is recommended that edge computing deploy, execute, control, schedule and coordinate its own applications, as well as coordinating and managing its infrastructure and services.

10 Security considerations

Security threats and requirements in [b-ITU-T X.1815] need to be considered for edge computing. In edge computing, [b-ITU-T X.1815] analyses the deployment scheme and typical application scenarios of its services, specifies the threats and requirements specific to them in IMT-2020 and thus establishes security capabilities for the operator to safeguard its applications.

Security aspects for consideration within the edge-computing environment, can be also addressed by security challenges for cloud computing as described in [b-ITU-T X.1601] and [b-ITU-T X.1644]. In particular, for security, [b-ITU-T X.1601] analyses threats and challenges, and describes capabilities that could mitigate the former and meet the latter. For security on a distributed cloud, [b-ITU-T X.1644] analyses threats and challenges and proposes guidelines against the former.

Relevant security requirements of [b-ITU-T Y.2701] and applicable ITU-T X-, ITU-T Y- and ITU-T M-series Recommendations need to be taken into consideration, including access control, authentication, data confidentiality, data retention policy, network security, data integrity, availability and privacy.

Appendix I

Analysis of edge computing for common characteristics from various application domains in industry

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

This appendix lists well-known application domains of edge computing in order to derive its common characteristics. The characteristics of edge computing are summarized in Table I.1.

Table I.1 – The applications of edge computing and its characteristics

Application domain	Usage of edge computing in application domain	Characteristics derived from usage of edge computing
Autonomous vehicle	• Safe-driving • Cooperative driver assistant • Cellular vehicle-to-everything communication • Smart city and IMT-2020, transportation system integration • Charging stations – Predict and plan • Personalized infotainment or reconfigurable cluster fleet management – A managed environment respecting policy and optimizations • Cloud-connected drones	• Low latency • Network connectivity • Minimal bandwidth consumption • Support of mobility
Industry and manufacture	• Real-time coordination of manufacturing processes using data from multiple machines, processes and systems • Edge-based conditional monitoring and filtering • Maintenance prediction through data analysis • Cloud use for flexible manufacturing processes (manufacturing as a service) • Eliminating downtime with industrial edge computing • Visibility into production cycles to optimize production • Machine learning and sensor fusion for industrial applications to provide an autonomous control platform • Robotics and automated factories	• Low latency • Proximity of user to data • Minimal bandwidth consumption • Support of mobility • Scalability of infrastructure
Smart city	• Traffic signal control • Work zone warnings • Transit bus traffic signal priority • School zone flashing beacons • Changeable message signs • Pay-as-you-drive insurance • Smart parking	• Low latency • Network connectivity • Proximity of user to data • Minimal bandwidth consumption • Support of mobility • Scalability of infrastructure

Table I.1 – The applications of edge computing and its characteristics

Application domain	Usage of edge computing in application domain	Characteristics derived from usage of edge computing
	• Controlling lighting to support energy consumption and adjusting lights based on city conditions • CCTV control and monitoring • Crowd management • Aiding police to solve crimes • Smart waste management	
Smart home	• Controlling energy and water consumption, home applications remotely with security system • Home automation using real-time data • Smart home appliances (washing machines, refrigerators, speakers, televisions and security systems) based on artificial intelligence (AI) • Room temperature, humidity and CO_2 measured among other things and displayed on mobile devices	• Network connectivity • Proximity of user to data • Minimal bandwidth consumption • Scalability of infrastructure
Smart retail	• Cold chain management for store efficiency and guaranteeing secure and hazard analysis and critical control point-compliant cooling systems • Occupancy management • Real-time supply-chain monitoring • Queue management • Omni-channel shopping – optimum personalized combination of all contact points • Intelligent logistics • Real-time product location tracking • Real-time weight sensor and product image recognition • Real-time employee facial and gesture recognition	• Low latency • Network connectivity • Proximity of user to data • Minimal bandwidth consumption
Smart agriculture	• Smart greenhouse • Control of soil erosion for agriculture • Water management for sustainable farming • Precision fertilization • Automated irrigation and compliance • Grain bin level monitoring and control • Herd health tracking • Smart weeding • Autonomous tractors • Agriculture drones	• Low latency • Network connectivity • Proximity of user to data • Support of mobility

Table I.1 – The applications of edge computing and its characteristics

Application domain	Usage of edge computing in application domain	Characteristics derived from usage of edge computing
Smart grid	• Demand prediction for smart power grids • Grid edge • Decentralizing energy delivery, enabling people to quickly connect to and disconnect from the larger grid and generate and deliver electricity locally • Distributed energy resources • Micro-grid • Energy-as-a-service • Multi-transport network platforms, distributed intelligence and low voltage distribution management	• Network connectivity • Scalability of infrastructure
Smart building	• Workplace distancing and contact tracing • Lift monitoring and predictive maintenance • Space monitoring • Usage-based cleaning • Indoor navigation • Air quality management • Parking lot monitoring • Local energy management • Integrated workplace management system • Better utilization of building assets and services, which occurs when building managers use edge AI to keep an eye on critical systems like elevators, fire alarms and fire extinguishers	• Low latency • Proximity of user to data • Minimal bandwidth consumption • Support of mobility • Scalability of infrastructure
Healthcare	• Fall detection • Remote patient monitoring • Mobile-to-mobile connectivity across or between medical devices, allowing the tracking of people and assets within the facility • Improved medical procedures and training through the use of emerging technologies such as augmented reality (AR) and virtual reality (VR) • Remote and virtual pain management using AR and VR • Remote or near real-time surgical assistance via the ability to layer multiple surgical techniques using AR • Telemedicine • Patient monitoring that leverages medical devices such as insulin pumps, smart lenses and pacemakers • Wearables and connected apps that track various health metrics	• Low latency • Network connectivity • Proximity of user to data • Minimal bandwidth consumption • Support of mobility • Scalability of infrastructure

Table I.1 – The applications of edge computing and its characteristics

Application domain	Usage of edge computing in application domain	Characteristics derived from usage of edge computing
	• Monitoring for occupant health conditions and regulation compliance including monitoring of elevated body temperature detection, social distancing monitoring and mask detection through correlation with vision-based sensors. (COVID-19)	
Transportation	• Routing emergency response vehicles around congested areas • Collecting data from embedded sensors, video cameras, crowd-sourced traffic information and other sources • Intelligent traffic systems, safer roads, directed parking, and frictionless toll and parking payments • Connected public transportation with urban mass transportation systems • AI-enabled traffic and weather detection, sensor-empowered interactive road lights, and smart pavements • Connected air travel –parking assistance on arrival at an airport, through check-in, boarding, and inflight entertainment to arrival and baggage claim • Preventing train and track failures, providing accurate real-time data and analytics, and facilitating pre-emptive maintenance • Shipping's digital transformation	• Low latency • Network connectivity • Proximity of user to data • Minimal bandwidth consumption • Support of mobility • Scalability of infrastructure
Gaming	• Deploying cloud games on demand from cloud to edge • Send a game control instruction from the cloud to multi-access edge computing (MEC), rendering the video on the MEC and sending the video stream to the terminal • Accelerating the rendering of images • Gaming traffic offloading • Optimizing network bandwidth usage and reduces latency	• Low latency • Network connectivity • Minimal bandwidth consumption • Scalability of infrastructure
Smart robot	• Product assembly to be carried out automatically • Putting systems intelligence in the cloud and simplifying robotics on the ground • Real-time capability of robot control functions • Smart robot applied healthcare system	• Low latency • Network connectivity • Scalability of infrastructure • Decentralized processing

Appendix II

General use case for edge computing

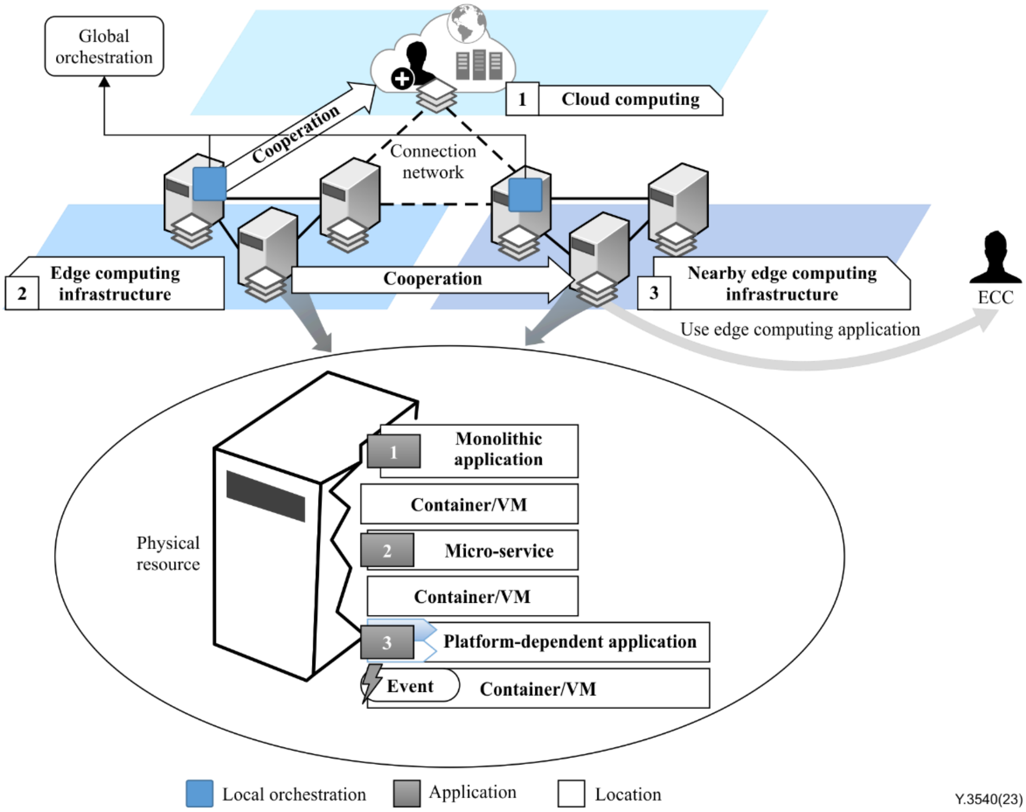
(This appendix does not form an integral part of this Recommendation.)

See Table II.1.

Table II.1 – General use case with capability of edge computing

Description	Edge computing utilizes some enterprise data centres or other technology services such as those of cloud computing and IoT. Furthermore, edge computing deploys its applications on its nearby infrastructure from an ECC such as a processor, storage or networks in a variety of distributed resource environments. Furthermore, by providing low-latency capability for data processing between an enterprise data centre including cloud computing and edge-computing infrastructure, low latency capability and cooperative capability selectively utilizes edge-computing infrastructure close to an ECC to provide faster data processing than the response of the centralized cloud infrastructure and ensure the stability of essential safety-critical real-time services. Currently, in order to overcome data processing and transmission delays due to centralization of large-scale data generation, edge computing has developed to provide its service related to that of cloud computing, which processes data at a location close to its application for data affinity capability. This use case describes an architecture that provides a high-speed network, which in edge computing connects multiple clusters as infrastructure based on a container, and thereby provides a service. The architecture and management are shown in Figure II.1. As shown in Figure II.1, the architecture involves cooperation between cloud computing, and distant and nearby edge-computing infrastructure. Cooperation requires network connectivity capability and constitutes storage to link the clusters and edge-computing infrastructure. For automated capability, edge computing orchestrates its application and service with one orchestrator that is global connected to another that is local. Three locations can accommodate not only cloud computing but also edge-computing infrastructure: • cloud computing; • distant edge-computing infrastructure; • nearby (nearest) edge-computing infrastructure. In addition, the edge-computing application supports three types of service as follows: • monolithic application service; • micro service; • specific platform-dependent application services (function as a service, PaaS, serverless computing platform, etc.) To support various edge-computing applications, the architecture is configured as shown in Figure II.1 and has the following characteristics. High-performance architecture for efficient cooperation between clusters: to increase container efficiency, high-speed storage (main memory, RAM, NVMe, etc.) is utilized to compose a high-performance container. High-speed network connectivity between clusters for cooperation: edge computing provides various capabilities for high-speed networks such as gateways for cluster master, tunnelling between clusters and high-speed routing for cluster for cooperation between clusters.

Table II.1 – General use case with capability of edge computing

	Orchestration technology for cooperation between clusters: orchestration of edge computing provides both a global and local orchestrator at each cluster. Through intelligent scheduling to optimize cooperation between cloud computing and edge-computing infrastructure, ECNs distribute an edge-computing application over nearby edge-computing infrastructure.
Role/sub-role	ECC
Figure (optional)	 VM: virtual machine **Figure II.1 – The general use case of edge computing**
Pre-conditions (optional)	
Post-conditions (optional)	
Derived high-level requirements	– [User proximity] – [Minimization of latency] – [Network connectivity for infrastructure] – [Data affinity] – [Infrastructure interface] – [Cooperation among applications] – [Resource provisioning] – [Orchestration]

Appendix III

Use cases for edge computing

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

This appendix describes edge-computing use cases applied to several application domains. See Tables III.1 to III.5.

Table III.1 – Use case of edge computing for healthcare service

Description	IoT devices, which include medical and healthcare devices such as an electrocardiogram, smart wristband, smart watch and blood pressure monitor, are used to monitor various biological indicators of the human body and generate health data. However, due to hardware limitations related to computing capability, storage capacity and power consumption, IoT devices cannot process such health data efficiently. Therefore, IoT users have to: (1) upload data to the cloud; (2) bring the data to the hospital to obtain the health analysis result. Both methods (1) and (2) meet the real-time requirement of applications that require analysis response immediately, and are inconvenient for user service, so there is a difficulty in early warning for potential problems of health. By executing tasks on nearby resources where the collected data is located, edge computing effectively reduces the latency of computing execution time, decreases data transmission, and protects data security and privacy. In addition, edge computing mainly utilizes biological indicators to execute AI inference related to diagnosis, executes the AI model deployed in a large-scale data centre such as cloud computing, and feeds the execution results back to an ECC or data centre. Devices in the IoT deliver monitored data through the local area network to nearby storage in edge-computing infrastructure, where a deep learning model for processing the data is deployed. For deep learning models for edge computing, there is a trade-off relationship between the accuracy of diagnosis and the complexity of performing the model. In other words, a deep learning model performed in edge computing has some limitations on accuracy and complexity. In addition, data privacy is considered for data transmission only within a local area network controlled by an ECC. Cloud computing in a large-scale data centre provides global, non-real-time, long-term, big data processing and analytics. Furthermore, cloud computing in large-scale data centre performs AI model training according to business requirements, historical data, real-time data and AI execution feedback. In addition, the AI model is also sent to edge-computing infrastructure for user proximity. An ECC obtains proper health monitoring and diagnosis from nearby edge-computing infrastructure and edge computing increases accuracy by passing large-scale data to cloud computing for more accurate inference. If an ECC authorizes hospitals and individual doctors to use its data, a more detailed and accurate diagnosis is obtained through edge computing. Edge computing using cloud computing service enables third party organizations to intelligently connect patients, clinicians, healthcare providers and insurance companies in a seamless and collaborative way, helping patients achieve a one-stop experience in nursing, healthcare and insurance services.

Table III.1 – Use case of edge computing for healthcare service

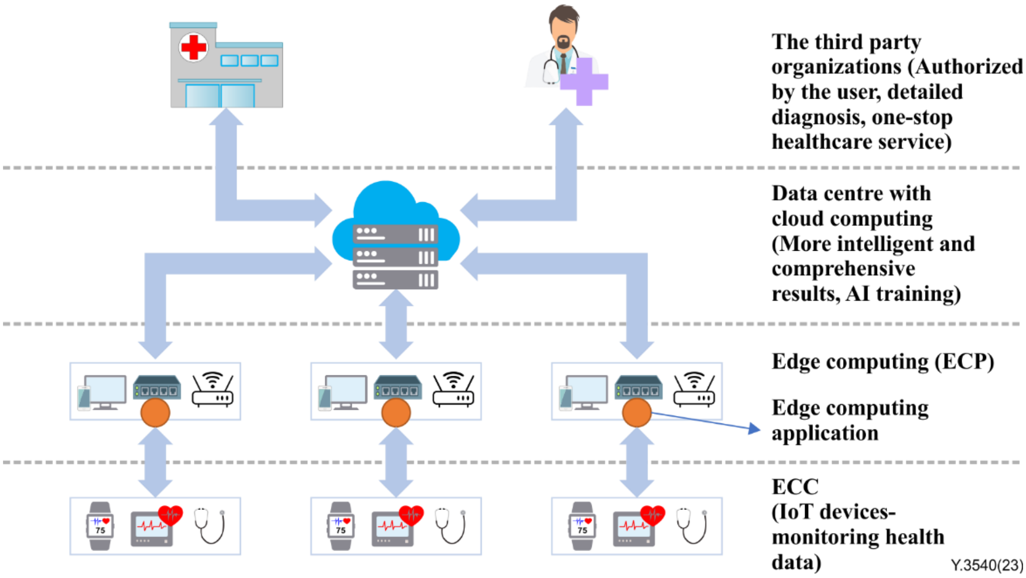
<table>
<tr><td></td><td>Resource management is a major issue in edge computing, where tasks are assigned to edge-computing infrastructure according to multiple factors such as energy, bandwidth consumption and latency. In order to set processing priorities and reduce total processing time based on data collected about a person's health condition, a job-scheduling mechanism is needed to determine whether multiple jobs need to be run on local devices, a remote cloud or both of the cluster type.</td></tr>
<tr><td>Role/sub-role</td><td>ECP</td></tr>
<tr><td>Figure (optional)</td><td>

Figure III.1 – The use case of edge computing for healthcare service</td></tr>
<tr><td>Pre-conditions (optional)</td><td></td></tr>
<tr><td>Post-conditions (optional)</td><td></td></tr>
<tr><td>Derived high-level requirements</td><td>– [Minimization of latency]
– [Orchestration]
– [Infrastructure interface]
– [Network connectivity for infrastructure]
– [Cooperation among applications]
– [Data privacy]
– [Resource provisioning]
– [Data affinity]
– [User proximity]</td></tr>
</table>

Table III.2 – Use case of edge computing for autonomous driving

Description	Autonomous vehicles are operated safely and reliably only when all data necessary for driving is analysed in real time. Furthermore, autonomous vehicles are operated with the installed sensor devices such as light detection and ranging, radio detection and ranging, camera, ultrasonic sensor and global positioning system. These sensors are mounted on the vehicle and interlock with the electronic control unit installed in the vehicle through, for example. a controller area network so that they react immediately to the situation.

Table III.2 – Use case of edge computing for autonomous driving

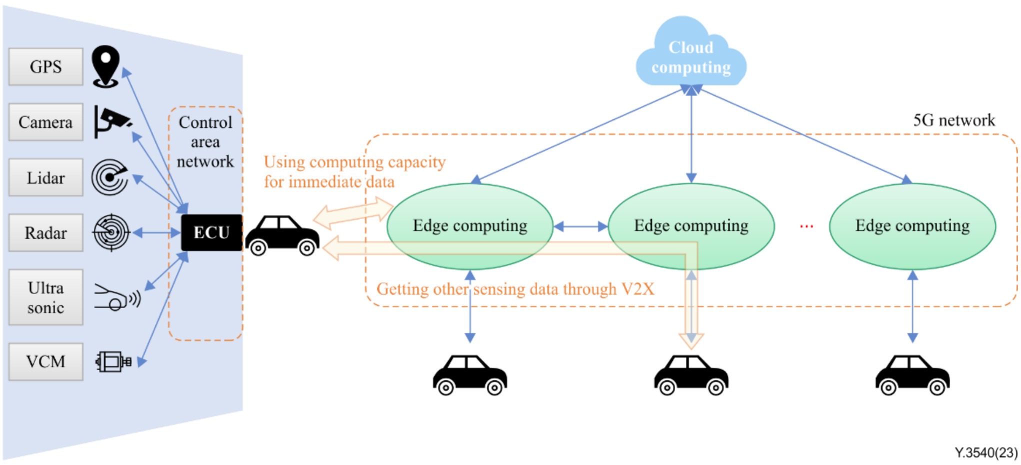
	Sensors installed in vehicles are accurate at close range, but have limitations in collecting data over long distances. So, a public IMT-2020 network is used outside the vehicle. In this case, if vehicle to everything (V2X) is applied, various data such as those related to vehicle operation, traffic conditions and warnings are received from other objects (including vehicles). Since data are shared between vehicles, implementation of cooperative autonomous driving is also possible. Even if a sensor with low performance is installed, V2X fills the gaps in low performance, so there is an advantage of lowering the price of the sensor installed in the car. A vehicle is equipped with a high-definition (HD) map in which surrounding environment information such as road, terrain, elevation, and curvature is implemented in three dimensions. The error range in an HD map is 10 cm, so a vehicle with HD provides information whose accuracy is 10 times higher than that with a general map. If the data from the sensors and the data used for V2X and the HD map are not delivered quickly, it leads to a major accident. Therefore, a faster real-time response is required to prevent an accident. These requirements are solved by introducing edge computing. It is possible to overcome the limited computing capability of the vehicle itself, and it is possible to use edge computing in its own infrastructure rather than cloud computing, which has excellent capability but a large latency to borrow other resources. Edge computing helps V2X communication relay efficiently and performs HD map real-time rendering. Of course, pre-processing is also possible before transmitting sensing data to cloud computing. Cloud computing handles data for the application of long-term rather than immediate data processing.
Role/sub-role	ECC, ECP
Figure (optional)	 **Figure III.2 – Use case of edge computing for autonomous driving**
Pre-conditions (optional)	
Post-conditions (optional)	
Derived high-level requirements	– [Minimization of latency] – [Orchestration] – [Network connectivity for infrastructure] – [Data affinity] – [User proximity] – [Mobility supports]

Table III.3 – Use case for smart home based on edge computing

Title	A use case for smart home based on edge computing
Description	A smart home is implemented using smart devices that converge sensors and communication technologies with domestichome appliances, home energy, home furniture and various IoT-based devices in the home based on IoT. By the real-time data transmission through the themsmart devices, an ECC in edge computing immediately and directly controls the smart devices. Such smart home technology allows ECCs to remotely access smart devices from the outside and use services such as systems for meter reading system, air conditioning or /heating system, lighting, and security. Most smart home services work with cloud computing to satisfy requirements forprovide a security network for smart home, unlimited home computing and storage resource services to satisfy smart home requirements. However, it is difficult for cloud computing-based smart home services to satisfy the increased requirements such as real-time and emergency handling, i.e. That is, in cloud computing, data collected from anthe IoT device is always transmitted to athe cloud-computing storage device of the cloud computing every time, which incurs high costs in terms of latency, storage capacity, bandwidth, and energy consumption. These requirements derive froms the concept of edge computing. In edge computing, resources such as a processor, storage or, network, etc. are available at the edge of the network close to the end smart device, and this arrangement reduces network latency. The smart home service scenario is shown in Figure III.3. Such aThe smart home includes smart devices that collect and transmit sensing data over thein IoT that require minimal human intervention. Smart devices detect and process data from the home environment, and transmit the data by connecting to edge-computing infrastructure at or near the physical location. In the edge computing, a smart home gateway is in charge of input or /output of data traffic from the smart home devices and to the cloud resources, respectively. The Edge computing performs calculatesion, communicatesion, and storesage for data instead of limited resources (memory, battery backup, processing capability, etc.) for smart devices. The Data analysis is also analysed inperformed at the edge-computing infrastructure and spend the less time is taken for decision-making than analysis inat cloud computing remotely. The resources ofor cloud- computing processes for large-capacity data that cannot be processed by edge-computing infrastructure collect pre-processed data transmitted by it. Useful and relevant information for families and communities is extracted from the data collected, whichand the cloud computing stores for a long time to help with heavy analysis. Authenticated smart home ECCs access smart home data through edge-computing infrastructure and resources in cloud computing. In edge computing, ECCs use anedge computing application and edge computing service immediately in situations that require quick response, such as emergenciesy situations, fires, and crime prevention, as well as simple control and monitoring of smart devices through the edge computing infrastructure.
Role/sub-role	ECC, ECP

Table III.3 – Use case for smart home based on edge computing

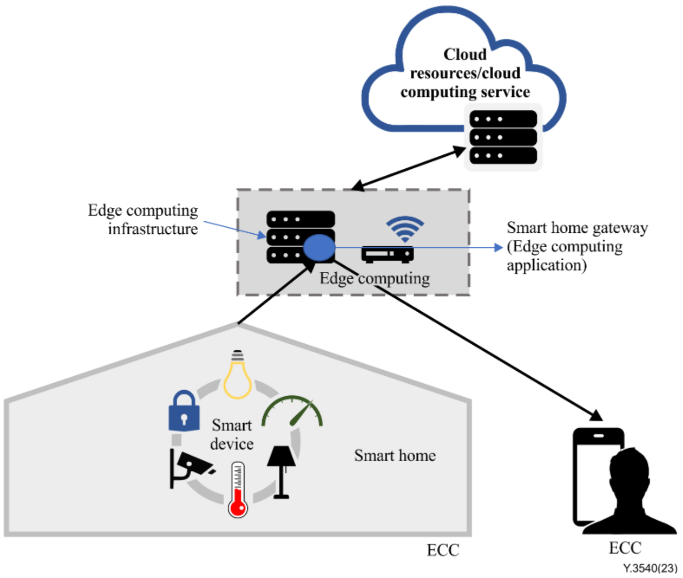
Figure (optional)	 **Figure III.3 – The Use case of edge computing for smart home**
Pre-conditions (optional)	
Post-conditions (optional)	
Derived high-level requirements	– [Minimization of latency] – [Orchestration] – [Data affinity] – [User Authentication].

Table III.4 – Use case for cooperation in edge computing

Description	In edge computing, infrastructure and a data centre including cloud computing in different locations are connected through a network, and pieces of infrastructure are connected to each other. Therefore, edge computing includes the capability to maximize efficiency for distributed infrastructure by workload offload and migration of its application. Figure III.4 shows a use case for cooperation in edge computing. In edge computing, a capability for cooperation includes workload offloading between data centres including cloud computing and infrastructure and migration between pieces of infrastructure. In edge computing, migration involves movement of an application and the related data. It is necessary to have a reliable and efficient method to capture the state of the application in real time in a transferable format to save and maintain the state of running applications such as snapshot. Offloading includes transfer of the role of resource-intensive computational tasks through monitoring of resource load and the network by global scheduling to a separate processor such as a hardware accelerator or an external device in a cloud data centre. For edge-computing, the migration and work offload of an application to be run on each piece of its infrastructure is driven by a capability enabler. This software

Table III.4 – Use case for cooperation in edge computing

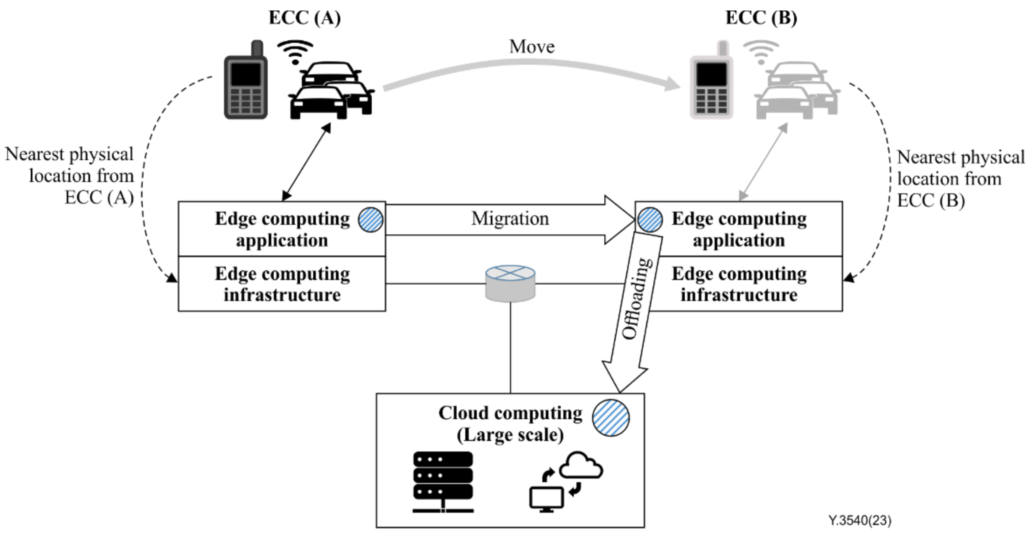
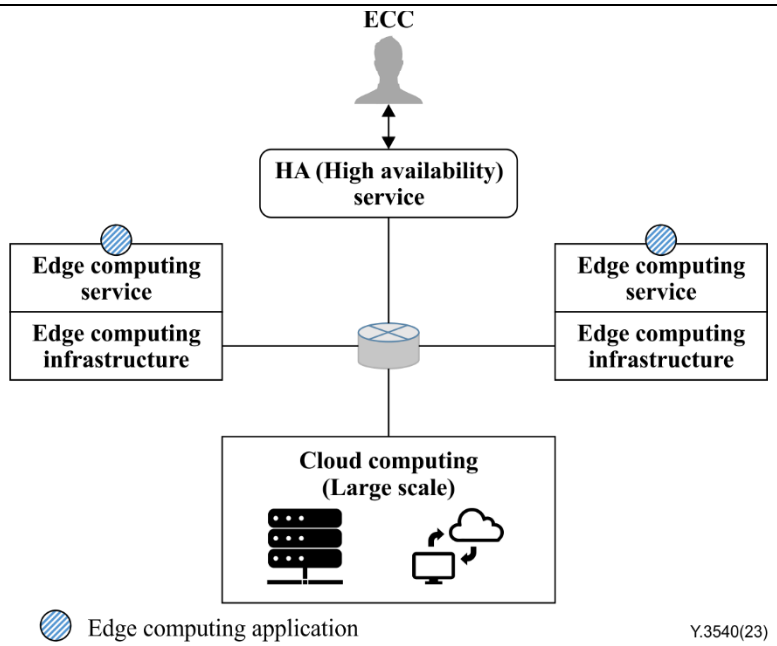
	administers and acts as executive agent for migration to provide edge-computing service to ECP. Furthermore,, in order to migrate its application and large-capacity data, edge computing provides a storage interface for data migration, and globally shared storage for high-speed network between pieces of edge-computing infrastructure.
Role/sub-role	ECC, ECP
Figure (optional)	 **Figure III.4 – Use case of edge computing for cooperation**
Pre-conditions (optional)	
Post-conditions (optional)	
Derived high-level requirements	– [Provision of service] – [Cooperation among applications]

Table III.5 – Use case for the network connectivity for edge computing

Description	This use case provides an example of network connectivity, which means network connection to enable continuous service to an ECC in edge computing. Network connectivity is presented as one capability in edge computing enabling continuous service according to the movement of an ECC and edge-computing application. Figure III.5 shows network connectivity capabilities of edge computing. As shown in Figure III.5, each piece of edge-computing infrastructure and data centre using cloud computing operate remotely and separately with different IP addresses. Edge-computing infrastructure is connected with the network at the request of an ECC due to cooperation and data affinity. If edge-computing infrastructure is already connected to the network, the service is available through that network. Otherwise, the network connection is only possible to connect with the public interface (IP address or URL). Also, if edge-computing infrastructure consists of clusters or includes a data centre with cloud computing, the access for an edge-computing application is protected and not exposed to the outside. Therefore, network connections utilize gateways between network layers, such as secure tunnelling-based networks using IP security or network

Table III.5 – Use case for the network connectivity for edge computing

	proxy-based routing rules using application layers in the transmission control protocol and native routing for awareness inside the cluster. Network connectivity capability supports the fail-over for an edge-computing application to prevent service interruption and to perform load balancing when a service is deployed. In Figure III.5, information about the service connection (such as IP, domain name or URL) is exposed to the edge-computing infrastructure. At this time, edge computing provides a high-availability service using HAProxy or another load balancer to support uninterrupted service according to service movement in both pieces of edge-computing infrastructure. The HAProxy server or load balancer checks the connection in real time and, if disconnected, performs a failover with a new connection using the service exposure. Edge computing provides a seamless network connection between ECC and edge-computing application. In edge computing, since its infrastructure (including cloud computing) needs to support newly created services or network connection requests from an ECC, gateways, routers and service exposures for network connectivity are implemented by an ECN using hardware or software resources and a service enables network connectivity on the infrastructure.
Role/sub-role	ECC, ECP
Figure (optional)	 **Figure III.5 – Use case of edge computing for network connectivity**
Pre-conditions (optional)	
Post-conditions (optional)	
Derived high-level requirements	– [Infrastructure interface] – [Network connectivity for infrastructure]

Bibliography

[b-ITU-T Q.1743] Recommendation ITU-T Q.1743 (2016), *IMT-Advanced references to Release 11 of LTE-Advanced evolved packet core network*.

[b-ITU-T X.1601] Recommendation ITU-T X.1601 (2015), *Security framework for cloud computing*.

[b-ITU-T X.1644] Recommendation ITU-T X.1644 (2023), *Security guidelines for distributed cloud*.

[b-ITU-T X.1815] Recommendation ITU-T X.1815 (2023), *Security guidelines and requirements for IMT-2020 edge computing services*.

[b-ITU-T Y.2060] Recommendation ITU-T Y.2060 (2012), *Overview of the Internet of things*.

[b-ITU-T Y.4100] Recommendation ITU-T Y.2066/Y.4100 (2014), *Common requirements of the Internet of things*.

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

[b-ITU-T Y.2701] Recommendation ITU-T Y.2701 (2007), *Security requirements for NGN release 1*.

[b-ITU-T Y.3123] Recommendation ITU-T Y.3123 (2023), *Framework of edge computing capability exposure for IMT-2020 networks and beyond*.

[b-ITU-T Y.3500] Recommendation ITU-T Y.3500 (2014), *Information technology – Cloud computing – Overview and vocabulary*.

[b-ITU-T Y.3502] Recommendation ITU-T Y.3502 (2014), *Information technology – Cloud computing – Reference architecture*.

[b-ISO 27729] International Standard ISO 27729:2012, *Information and documentation – International standard name identifier (ISNI)*.

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