International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Q.5001
(10/2018)

SERIES Q: SWITCHING AND SIGNALLING, AND ASSOCIATED MEASUREMENTS AND TESTS

Signalling requirements and protocols for IMT-2020 – Signalling requirements and architecture of IMT-2020

Signalling requirements and architecture of intelligent edge computing

Recommendation ITU-T Q.5001

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
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2020	Q.5000–Q.5049
Signalling requirements and architecture of IMT-2020	**Q.5000–Q.5019**
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.5001

Signalling requirements and architecture of intelligent edge computing

Summary

A large volume of data have been generated from the use of various types of smart things. The related smart services have been working based on cloud systems. However, various issues has occurred as a result of the network bottleneck between terminals and a cloud system (e.g., data loss, network delay, etc.). An edge computing technology between the user equipment and a cloud server system is envisaged to solve these problems. In addition, applying the intelligent data processing functions by providing artificial intelligence (AI) technologies will provide enhanced networking capabilities for new emerging services and applications.

Regarding these emerging environments, Recommendation ITU-T Q.5001 defines the intelligent edge computing (IEC). It is applicable to collect, store, and process data reliably in the intelligent edge computing, especially to support mission critical services. Thus, the main functionality of intelligent edge computing is collecting, processing, analysing the data and providing the values based on intelligent data processing.

This Recommendation specifies use cases, signalling requirements and an architecture of intelligent edge computing.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Q.5001	2018-10-14	11	11.1002/1000/13701

Keywords

Architecture, edge computing, edge service, requirement, use case.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

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 received notice of intellectual property, protected by patents, 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 TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	1
3.1	Terms defined elsewhere	1
3.2	Terms defined in this Recommendation	1
4	Abbreviations and acronyms	1
5	Conventions	2
6	Overview	3
7	Signalling architecture	4
7.1	IEC signalling architecture model	4
7.2	Reference point	6
8	Signalling requirements of reference points	8
8.1	Signalling requirements for IEC with edge networking capability	8
8.2	Signalling requirements for IEC with data processing capability	9
9	High-level signalling protocol procedures	10
9.1	Signalling protocol procedures for networking functions	10
9.2	Signalling protocol procedures for intelligent data processing	15
9.3	Characteristics of signalling messages	17
10	Deployments	19
11	Security considerations	20
Appendix I – Use cases of intelligent edge computing		21
I.1	Mobile video delivery optimization using throughput guidance for Transmission Control Protocol (TCP)	21
I.2	Active device location tracking	21
I.3	Bandwidth allocation manager for applications	21
I.4	Video caching, compression and analytics service chaining	22
I.5	Application computation off-loading	22
I.6	Data path off-loading	22
I.7	Intelligent data processing and filtering	23
I.8	Smart construction monitoring system using machine learning techniques	23
Appendix II – Related works of intelligent edge computing		24
Bibliography		25

Recommendation ITU-T Q.5001

Signalling requirements and architecture of intelligent edge computing

1 Scope

This Recommendation specifies signalling requirements and an architecture of intelligent edge computing to provide intelligence to the edge network for efficient data processing within the network. It describes the following details:

– overview of intelligent edge computing;
– signalling architecture;
– signalling requirements;
– high-level signalling protocol procedures;
– use cases of intelligent edge computing; and
– related works of intelligent edge computing.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term 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.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 intelligent edge computing: Intelligent edge computing is a network architecture concept that enables edge networking and data processing capabilities for edge analytics by applying artificial intelligence technologies.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence
AIaaS Artificial Intelligence Infra as a Service
API Application Programmable Interface
CCTV Closed Circuit Television
DL Deep Learning
DQ Data Quality
EAB Edge Accelerated Browser
EGE Edge Gateway Entity
EME Edge Identity Management Entity

eNB	evolved Node Base station
ENE	Edge Networking Entity
ETL	Extract, Transform, and Load
GPS	Global Positioning System
HTTP	Hyper Text Transfer Protocol
IaaS	Infra as a Service
ICE	Intelligent Computing Entity
IEC	Intelligent Edge Computing
IoT	Internet of Things
IT	Information Technology
LTE	Long Term Evolution
MCS	Mobile Crowd Sensing
MEC	Mobile Edge Computing
ML	Machine Learning
QoE	Quality of Experience
RL	Reinforcement Learning
TCP	Transmission Control Protocol
TE	Terminal Entity
TLV	Type Length Value
WLAN	Wireless Local Area Network

5 Conventions

In this Recommendation:

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

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

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

6 Overview

Intelligent edge computing is a network architecture concept that enables edge networking and data processing capabilities for edge analytics by applying artificial intelligence (AI) technology.

The volume of data for AI is increasingly exploding. If a network equipment provides data from a terminal entity (TE) to cloud computing, data may get lost or network delay occurs because of network bottleneck. Therefore, there is the need for the equipment to collect data reliably and to provide valuable data to the cloud computing for big data analytics. Moreover, through machine learning (ML), there is need for the capability to analyse the data and to respond promptly. Mainly mission critical services (real-time and highly-reliable services) can be offered through processing

and analysing data generated from this equipment. Through IEC, it is possible to provide reliable and prompt mission critical service such as nuclear power plant alarm and intelligent traffic light control services. There are two important aspects: autonomous network control for an intelligent edge networking capability, and ambient intelligence analytics for intelligent data processing capability. IEC can consist of the four main entities depicted in Figure 6-1; edge networking entity (ENE), intelligent computing entity (ICE), edge gateway entity (EGE) and edge identity management entity (EME). Basically, IEC could be deployed in a location between terminal entity (TE), which includes a computing device such as smart phone, computer, and Internet of things (IoT) devices, and big data analytics at cloud computing. These entities can also be deployed either as a physical equipment or a virtual equipment, but this specification is out of scope of this Recommendation.

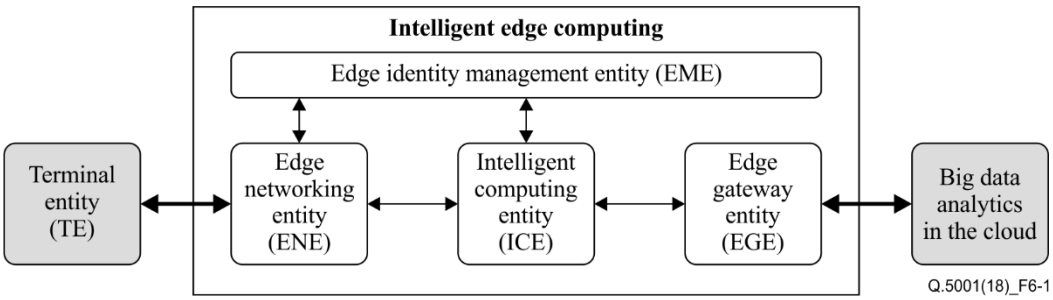


Figure 6-1 – Intelligent edge computing concept

– Edge networking entity (ENE): This entity provides connectivity to TEs using heterogeneous wireless technologies in resource constrained environment. Thus, it should be deployed for various wireless technologies as well as protocols between TEs and IEC.

– Intelligent computing entity (ICE): This entity provides edge analytics of AI service on itself or other analytics, such as big data analytics on cloud computing. Thus, it performs data analysis functions based on gathered information. In addition to the edge analytics, it should control TEs through the result of the analytics.

– Edge gateway entity (EGE): This entity provides an interworking function to outside entities including other IECs and big data analytics on cloud computing. Thus, it can function as a gateway function.

– Edge identity management entity (EME): This entity provides a management function that stores the identity of entities such as TE, ENE, ICE, and EGE, including data names as an identity. It also maps these identities to metadata such as ICE locations, and EGE locations. It could therefore be involved in mobility management, such as TE's mobility.

IEC should consider the following two important capabilities:

– Intelligent edge networking capability: In order to support mobility, real-time communication, reliable communication, scalability, constrained environment and easy deployment, IEC should consider networking technologies, such as applying various wireless technologies, backward compatibility with existing networks, and composing networking dynamically.

– Intelligent data processing capability: In order to support edge analytics, IEC should increase data utilization through intelligent data processing, including data collection, dynamic storage, and real-time trust data process. In particular, analysis models should be updated periodically by a big data server.

7 Signalling architecture

In this clause, it defines the signalling architecture which is composed of basic functions to enable two IEC capabilities. The IEC enables an edge analytics to achieve an intelligent edge networking and intelligent data processing capability.

7.1 IEC signalling architecture model

Figure 7-1 represents the overall IEC signalling architecture. The architecture consists of many functional blocks deployed in each IEC entity, such as ENE, ICE, EGE and EME. As all the functions could be classified under two IEC capabilities, these functions and interfaces are explained by separated architectures.

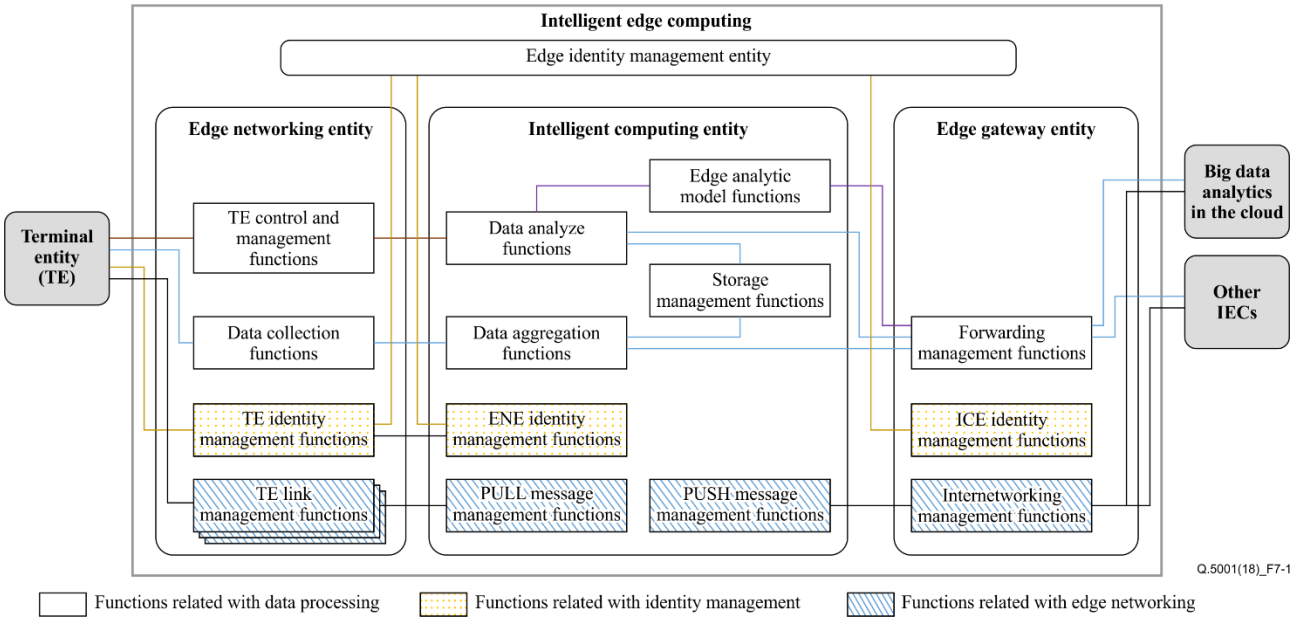
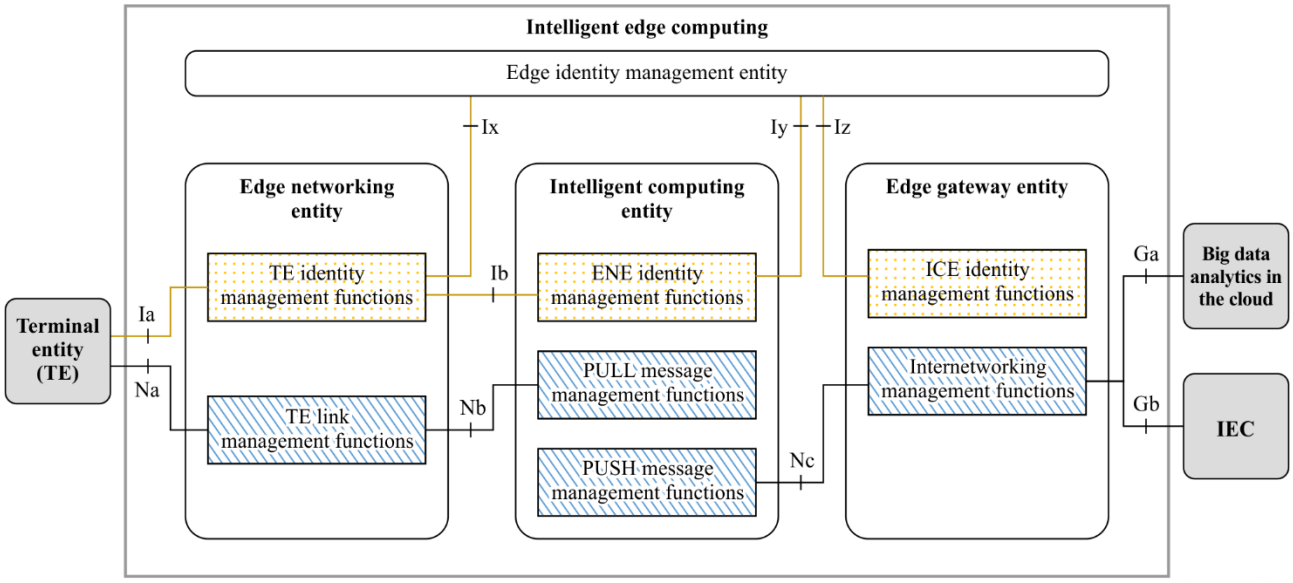


Figure 7-1 – Overall IEC signalling architecture

7.1.1 Signalling architecture with edge networking capability

To support intelligent edge networking in IEC, a new management entity is defined. This entity is called edge identity management entity (EME). EME manages and maps the identities of all the entities to metadata such as ICE locations and EGE locations.



Nx – Identifier for edge networking reference points
Ix – Identifier for identity management reference points
Gx – Identifier for edge gateway reference points

Figure 7-2 – Signalling architecture with edge networking capability

Figure 7-2 depicts the IEC signalling architecture to compose edge networking between TE and big data analytics at cloud.

The following are the functions of an ENE that provide TE connectivity and identity management:

– TE link management functions establish a connectivity between TE and IEC through low-layer protocols such as a layer 2 protocol. Thus, these functions allow the adaptation layer to provide various wireless network technologies and protocols.

– TE identity management functions manage the identity of TEs as well as data identity generated from the TE through interworking with EME.

The following are the functions of an ICE that enable message system and data processing between IEC entities:

– PULL and PUSH message management functions take request and reply system as a PULL and PUSH type message system for collecting data.

– ENE identity management functions manage ENE's identity to register it to EME.

The following are the functions of an EGE that provide interworking with other systems:

– Interworking management functions manage an interface between IEC and cloud system, or outside other IECs.

– ICE identity management functions manage their own ICE identity to interwork with others.

7.1.2 Signalling architecture with data processing capability

To support intelligence data processing in IEC, raw data generated from terminal entities (TE) should be processed through different processing phases in each entity of the IEC. First, raw data could be collected, and text data could be online or flow based processed by processing functions, including pre-processing such as extract, transform, load (ETL), and machine learning for edge analytics.

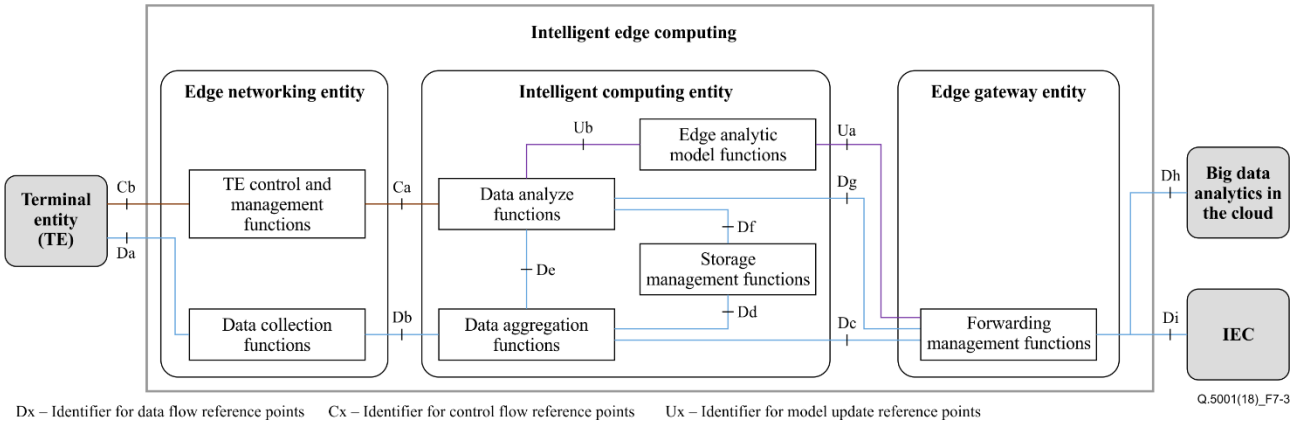


Figure 7-3 – Signalling architecture with data processing capability

Figure 7-3 illustrates the IEC signalling architecture to support an intelligent edge analytics capability including dynamic analytic model update functions through interworking between the IEC and a big data analysis server on cloud computing.

The following are the functions of an ENE to support collecting data and TE control management:

– Data collection functions collect raw data from TEs.

– TE control and management functions handle TE's operation and data management according to an action message by result of edge analytics.

Following are the functions of an ICE to enable data analysis in applying AI service:

– Data aggregation functions aggregates raw data from TEs through the ENE and it can support a message queue system to enable data stream processing.
– Data analysis functions apply a machine learning (ML) model to aggregated data by either stream processing or batch processing. According to the result of edge analytics, three categorized actions should be performed:
– The ML model is required to update dynamically by interworking with a big data analysis server on cloud computing;
– Storage management functions stores the result of edge analytics and aggregation data through the data aggregation functions.
– Edge analytic model functions first request an ML model from cloud computing and sends the reply to the serving model. Additionally, the ML model should be updated periodically.

The following EGE function supports forwarding the result of raw data or data of analytics to others:

– Forwarding management functions forward data to other systems, such as a big data analysis server on cloud computing and other IECs.

7.2 Reference point

7.2.1 Reference point Nx

These reference points denote signalling interfaces to establish connectivity between functions in each of the entities. Most of the control messages for composing network are required to carry out exchanges through the following interfaces:

– Na denotes an interface between a TE and the TE link management functions to establish connectivity.
– Nb denotes an interface between the TE link management functions and the PULL and PUSH message management functions to exchange a request and reply messages.
– Nc denotes an interface between the PULL and PUSH message management functions and the interworking management functions to establish connectivity.

7.2.2 Reference point Gx

These reference points denote signalling interfaces to establish connectivity between the IEC and outside system, such as a big data analytics on cloud computing and other IECs. Most of the control messages for interworking are required to carry out exchanges through the following interfaces:

– Ga denotes an interface between the interworking management functions and outside big data analytics at cloud.
– Gb denotes an interface between the interworking management functions and other IECs.

7.2.3 Reference point Ix

These reference points denote signalling interfaces to exchange identity registration or lookup messages toward the EME. Most of the control messages for registration of the identity are required to carry out an exchange through following interfaces:

– Ia denotes an interface between a TE and the TE identity management functions to notice attachment event.
– Ib denotes an interface between a TE identity management functions and the ENE identity management functions to register the TE's identity.
– Ix denotes an interface between a TE identity management functions and the EME to resolve identities of entities or data.
– Iy denotes an interface between the ENE identity management functions and the EME to register ENE identity to the ENE.

– Iz denotes an interface between the ICE identity management functions to resolve ICE identity.

7.2.4 Reference point Dx

These reference points denote signalling interfaces to forward data to build a data pipeline from a TEs to the outside systems. In the IEC, raw data should be transformed to information including meaningful metadata such as average, time stamp, location, etc., by edge analytics. Most of the data are required to be channelled through the following interfaces:

– Da denotes an interface between a TE and the data collection functions to collect raw data.

– Db denotes an interface between the data collection functions and the data aggregation functions to aggregate data.

– Dc denotes an interface between the data aggregation functions and the forwarding management functions to directly forward aggregate data toward outside systems.

– Dd denotes an interface between the data aggregation functions and the storage management functions to store aggregated data which were provided to the data analyse functions.

– De denotes an interface between the data aggregation functions and the data analyse functions to forward aggregate data directly to analyse them for online analysis.

– Df denotes an interface between the data analyse functions and storage functions to store analysed data or to be provided as aggregation data from database.

– Dg denotes an interface between the data analyse functions and the forwarding management function to forward the result of the analysis.

– Dh denotes an interface between the forwarding management functions and an external big data analytics cloud.

– Di denotes an interface between the forwarding management functions and other IECs.

7.2.5 Reference point Cx

These reference points denote signalling interfaces to control TEs according to the result of edge analytics. Most of the control messages are required to be channelled through the following interfaces:

– Ca denotes an interface between the data analyse functions and the TE control and management functions to control the TE's actions.

– Cb denotes an interface between the TE control and management functions and a TE to control the TE’s actions directly.

7.2.6 Reference point Ux

These reference points denote signalling interfaces to request or to reply to the ML model from a big data analytics at cloud computing. Most of the control messages are required to send requests to the ML model periodically through the following interfaces:

– Ua denotes an interface between the edge analytic model functions and the forwarding management functions. It is used to send requests to the serving ML model to a big data analytics on cloud computing.

– Ub denotes an interface between the data analysis functions and the edge analytic model functions to serve the ML model.

8 Signalling requirements of reference points

This clause describes signalling requirements to be considered in developing two capabilities for IEC architecture.

8.1 Signalling requirements for IEC with edge networking capability

8.1.1 Requirements for reference point Nx for edge networking functions

- General
  - Message exchanges are recommended to use a simple message format and to be carried out in few message handshakes
- Reference point Na
  - Connectivity is required to consider constraint environments

NOTE – Constraint environments include device constraints such as CPU, memory and other resource restrictions, and network constraints such as low power wide area network, etc.

  - TE's attachment is recommended to be frequently considered.
  - Movement of TE are required to be considered for data consistency.
  - It is recommended to provide low-latency connectivity to support mission critical service.
- Reference point Nb
  - It is recommended to provide low-latency connectivity to support internal push and pull communication.
- Reference point Nc
  - It is recommended to provide low-latency connectivity to support external push and pull communication.

8.1.2 Requirements for reference point Ix for identity management functions

- General
  - It is required to establish secure connections.
  - The control messages can optionally modify networking state to compose edge networking dynamically.
- Reference point Ia
  - The messages are recommended to be replied to within a limited timeframe to support frequent attachment of TEs.
- Reference point Ib
  - The messages are recommended to be forwarded to default ICE's function.
- Reference point Ix
  - TE or data identities are required to be securely registered.
- Reference point Iy
  - All information related to identities inside the IEC are required to be securely updated.
- Reference point Iz
  - Control messages from outside of the IEC can optionally use identity management function.

8.1.3 Requirements for reference points Gx for networking functions to interwork with external systems

- General
  - It is required to support various networking protocols to interwork with different networking protocols.
  - It can optionally build secure communication channel toward external systems.

– Reference point Ga
  – It is recommended to use standardization protocols.
– Reference point Gb
  – It can optionally establish a connection using low-layer protocols.

NOTE – For interworking between IECs within one administrator domain, non-secure channel can be established.

8.2 Signalling requirements for IEC with data processing capability

8.2.1 Requirements for reference points Dx for data flow functions

– General
  – Data are required to be identified by the identity management.
  – Data are recommended to be sent to target functions in accordance with pipeline operations.
– Reference point Da
  – Serialized data generated from TEs are recommended to be collected by sequential order.
  – Split data generated from TEs are recommended to be aggregated in the common message queue or system at the ICE.
– Reference point Db
  – Serialized data are recommended to be forwarded by sequential order.
  – Time-sensitive data are required to be forwarded within a given time frame.
– Reference point Dc
  – Raw data can optionally be directly forwarded to an outside system.
– Reference point Dd
  – Aggregated data can optionally be stored to light-weight database in advance.
– Reference point De
  – Time-sensitive data are required to be processed within a given time frame.
  – Serialized data are recommended to be forwarded for stream processing directly.
– Reference point Df
  – The result of edge analytics is recommended to be stored at light-weight database.
  – Data from storage are recommended to be provided as a batch file.
– Reference point Dg
  – Analysed data are recommended to be forwarded securely and steadily.
– Reference point Dh
  – Data are recommended to be securely forwarded to external servers.
– Reference point Di
  – Analysed data can optionally be forwarded to other IECs for other processing.

8.2.2 Requirements for reference point Cx for control functions to TE

– General
  – The control message is recommended to be forwarded by low-latency.
  – The communication is recommended to support a simple message exchange.
– Reference point Ca

  - The control messages are recommended to order operation in time for a mission-critical service.
- Reference point Cb
  - The control messages are recommended to be securely communicated.
  - The control messages are recommended to be forwarded in low-latency.

8.2.3 Requirements for reference point Ux for model management functions

- General
  - It is recommended to update the ML model periodically.
- Reference point Ua
  - It is recommended that an initial ML model be requested from an outside big data analytic system according to the service profile.
- Reference point Ub
  - It is recommended to immediately apply the external ML model.

9 High-level signalling protocol procedures

This clause describes and illustrates signalling flows to support intelligent edge networking capability and intelligent data processing capability.

9.1 Signalling protocol procedures for networking functions

Signalling protocol procedure with intelligent networking capability are classified into four procedures including the establishment of the TE's connectivity, gathering and obtaining data from a big data analytics, and TE mobility support as well as the establishment of inter-IEC depicted in Figures 9-1 to 9-4 respectively.

9.1.1 Procedures for the establishment of TE's connectivity

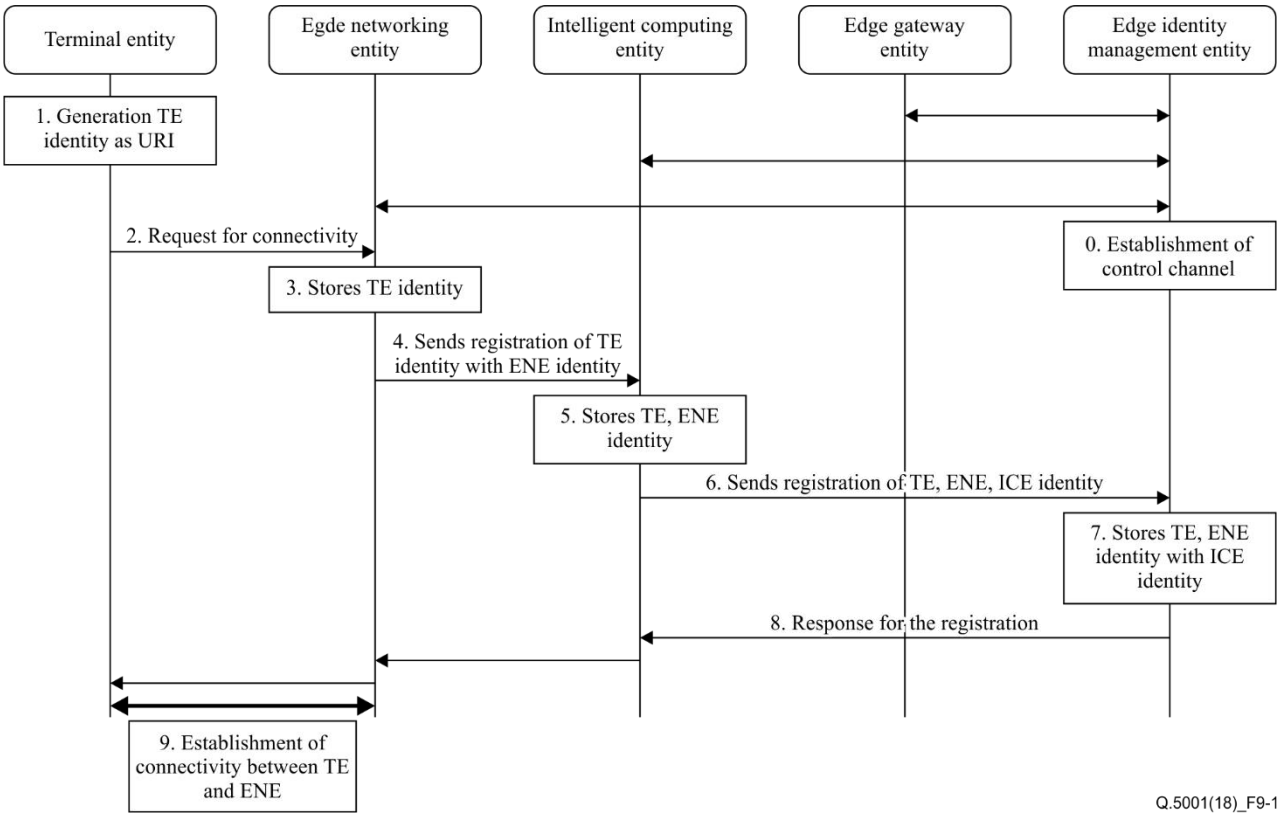


Figure 9-1 – Procedures for establishment of TE's connectivity

0. IEC entities such as ENE, ICE, and EGE could be connected to EME as a control channel by bootstrapping procedure.

   NOTE 1 – Bootstrapping procedure could be out of scope.

1. TE can generate its own identity by using hierarchical name such as URI type.

   NOTE 2 – TE identity could be local unique or global unique.

   NOTE 3 – TE identity generation procedures could be out of scope.

2. TE sends a request for connectivity with ENE.
3. ENE can manage served TE information by maintaining a table which includes served TE identities.
4. ENE sends a registration message to ICE to notify the TE identity and serving ENE identity.
5. ICE stores new TE identity with serving ENE identity at local cache.
6. ICE sends a registration message to EME to map their identities with the ICE identity.
7. EME stores their identities, thus there should be a mapping between TE, ENE and ICE identities.

   NOTE 4 – EME dynamically maps entity identities and various meta information such as locations and profiles.

8. EME sends a response message for the registration to the TE.
9. Connection between the TE and the ENE is established.

9.1.2 Procedures for gathering and obtaining data from big data analytics

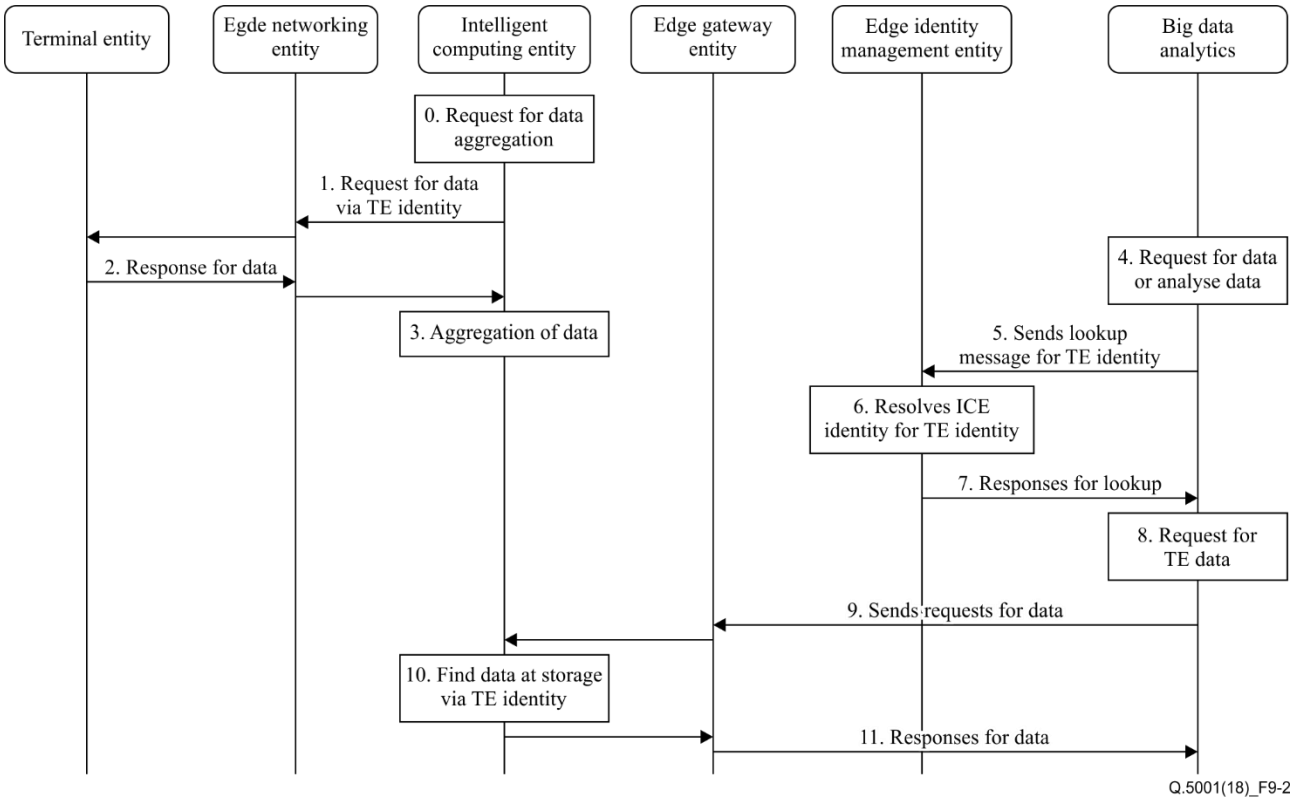


Figure 9-2 – Procedures for gathering and obtaining data from big data analytics

0. ICE knows served TE identity as well as data according to procedure for connection establishment, thus it sends request message for gathering data from the TE.

   NOTE 1 – A service profile could be registered by using TE identity registration procedure.

   NOTE 2 – A service profile includes data type, data identity, action type, etc.

   NOTE 3 – A service profile related specification is outside the scope of this Recommendation.

1. ICE sends a request message to the TE to collect data by using TE identity.
2. TE replies to the request from ICE, and triggers data generation.
3. ICE aggregates raw data sequentially via aggregation functions or storage functions.
4. A big data analytic server on the cloud continuously wants to obtain data generated from the TE.
5. The server sends a lookup message to resolve ICE identity that corresponds to TE identity.
6. EME locates TE identity to resolve the ICE identity.

   NOTE 4 – EME could be deployed in hierarchical servers to support scalability.

   NOTE 5 – EME structure is outside the scope of this Recommendation.

7. EME responds to the lookup request and the message includes the corresponding ICE identity.
8. The server wants to send a request for obtaining data to the resolved ICE.
9. The server sends a request message for obtaining data to the ICE.
10. ICE resolves the request by checking TE identity at storage.
11. ICE replies to the request by attaching data continuously.

9.1.3 Procedures for TE mobility support

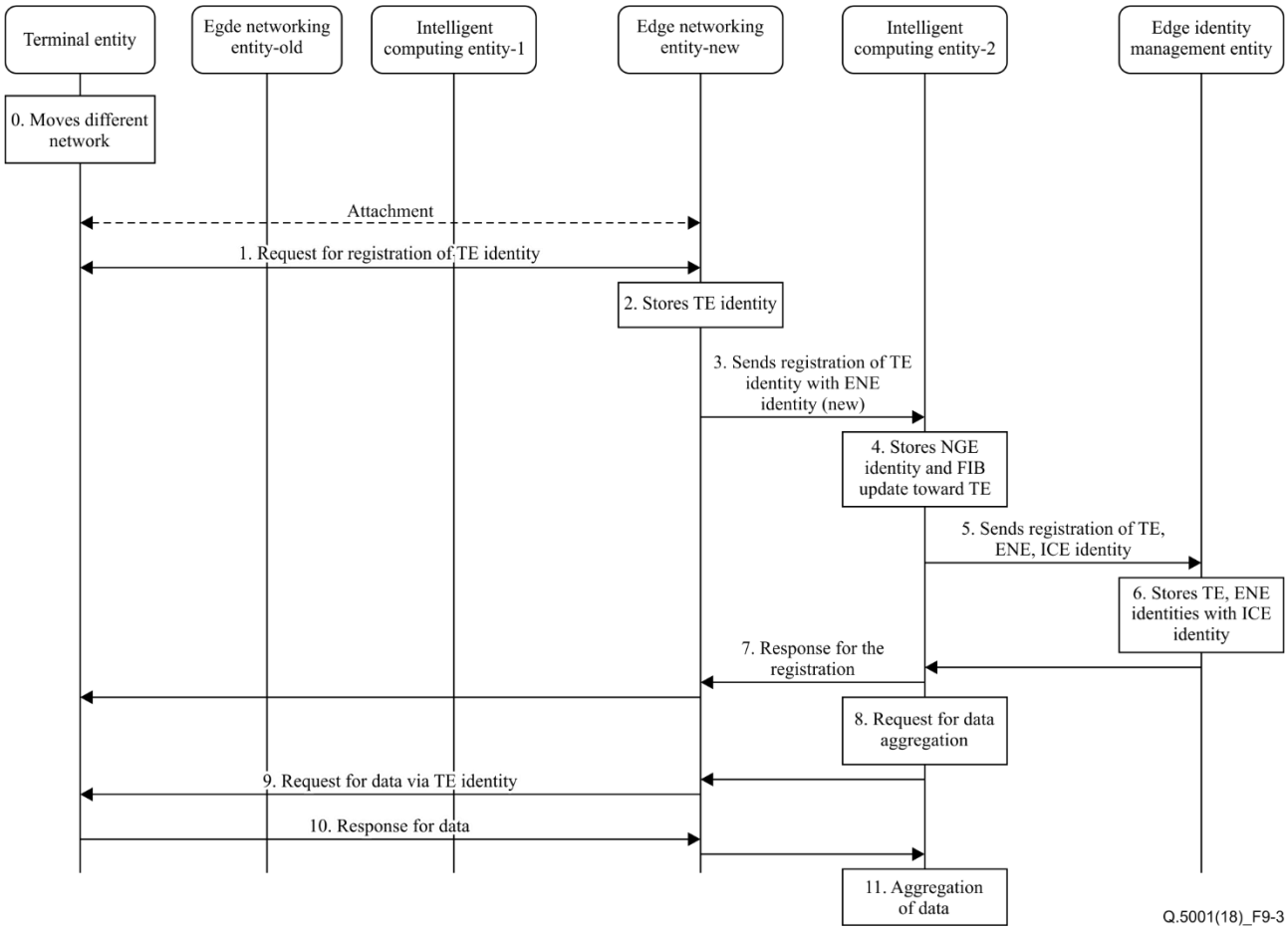


Figure 9-3 – Procedures for TE mobility support

0. A TE moves to a different network domain.

   NOTE – Attachment procedure could be out of scope.

1. TE triggers a registration request to the new ENE including TE identity.
2. ENE stores served TE identity.
3. ENE sends a registration message to the new ICE to notify the TE identity and the serving ENE identity.
4. ICE stores the TE identity and the serving ENE identity.
5. ICE sends a registration message to EME to map TE identities with new ICE identities.
6. EME stores their identities, thus it should be mapping between TE, ENE and ICE identities.
7. EME sends a response message to the registration request to the TE.
8. ICE wants to aggregate data continuously.
9. ICE sends the request to gather data from the TE.
10. TE replies to the request from the ICE, and re-triggers data generation.
11. ICE aggregates raw data sequentially to store them.

9.1.4 Procedures for inter-IEC communication

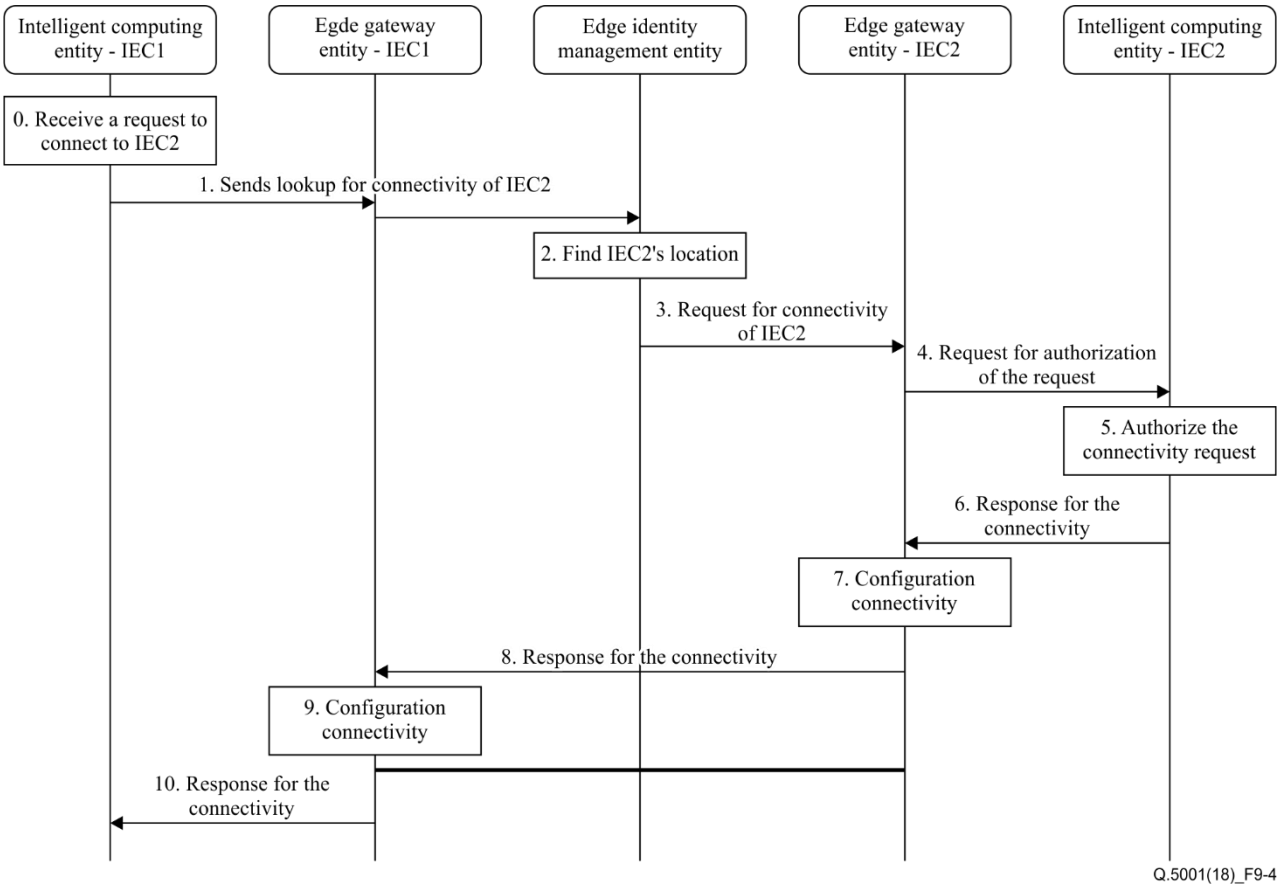


Figure 9-4 – Procedures for interworking between IEC1 and IEC2

0. ICE can receive a request for the establishment of other IECs.

   NOTE 1 – A service provider or an administrator can request the interworking between different IECs.

   NOTE 2 – The request can be achieved in out of band for communication, so it is outside the scope of this Recommendation.

1. ICE1 sends the request message to EME through EGE.

   NOTE 3 – The request message can include a lookup message to find corresponding IEC as IEC2.

2. EME finds IEC2's location as an EGE2's identity.
3. ENE sends a connection request message to EGE2 requesting the location of ICE2.
4. EGE2 sends a request authentication for the connectivity request to ICE2 as a representative of IEC2.
5. ICE2 authorizes the request to establish connection between IEC1 and IEC2 via EGE1 and EGE2.
6. ICEs sends a response message for the request to EGE2.
7. EGE2 prepares to configure the establishment of connectivity with EGE1.
8. EGE2 forwards the response message to EGE1.
9. EGE1 establishes the connection between EGE1 and EGE2.

   NOTE 4 – If a secure communication channel is required between IECs, then more communication procedures can be achieved, but this is outside the scope of this Recommendation.

10. EGE1 forwards the response message including the result of the connection.

9.2 Signalling protocol procedures for intelligent data processing

Signalling protocol procedure with intelligent data processing capability are classified into three procedures, including data pipeline from collecting to forwarding to target entity or systems, control TE according to the result of the analysis, and edge analytic model update dynamically.

9.2.1 Procedures for data lifetime from collecting to forwarding data target entity

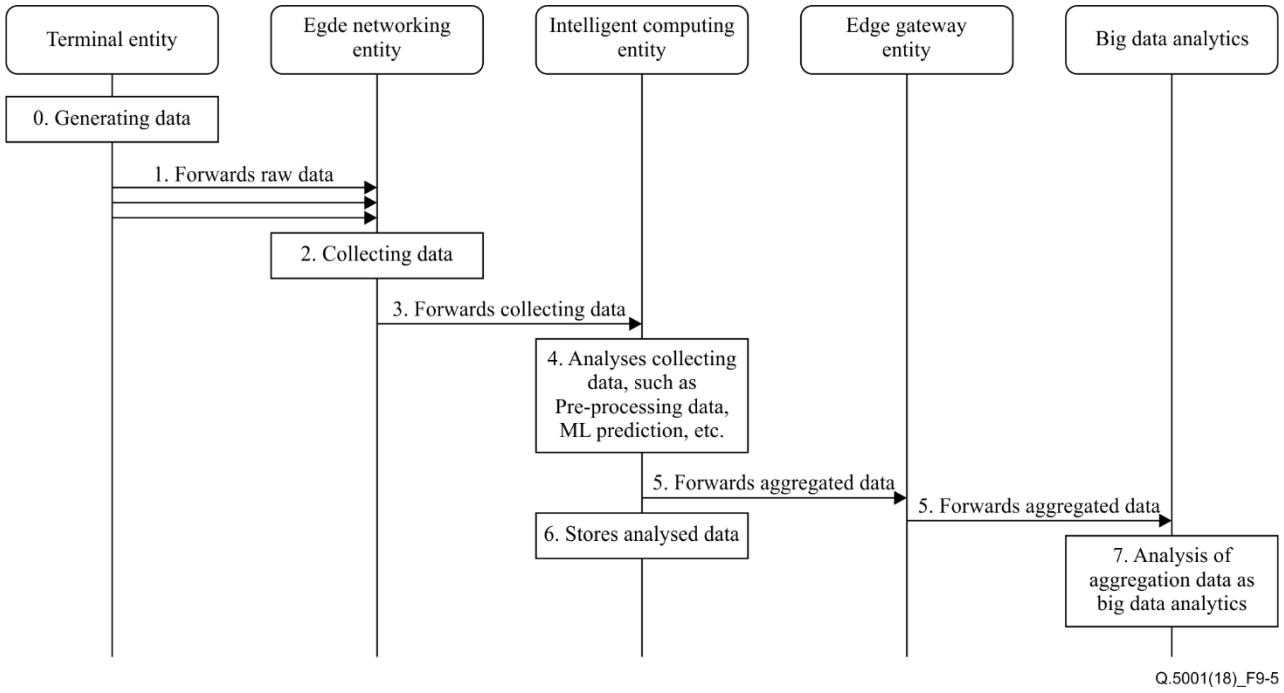


Figure 9-5 – Procedures for data lifetime

0. TE can generate raw data (or single unit data) via some triggering event.

   NOTE 1 – Some triggering can be done by TE, big data analytics, ICE and so on.

1. TE forwards raw data to ENE.

   NOTE 2 – Partitioned raw data can be collected by ENEs via connectivity between TE and ENE.

2. ENE collects partitioned data.

3. ENE forwards collecting data to ICE.

4. ICE analyses aggregation data.

   NOTE 3 – Aggregation data can be either stored in a storage or analysed directly as stream data processing.

   NOTE 4 – Aggregation data can be first processed in a pre-processing phase such as data quality (DQ) and extract, transform and load (ETL). Second, normalized data can be processed by using the AI model. For instance, in order to predict future event, ML prediction model can be applied to the edge analytics.

5. As a result of edge analytics, ICE forwards analysed data to the big data analytics server on cloud computing.

   NOTE 5 – Similar to Note 4, according to the result of ML prediction, forwarding data can be controlled to reduce the traffic load between ICE and the cloud server by using a type of video quality adaptation function, such as video transcoding function.

6. Edge analysed data are stored via storage function.

7. Big data analysis of edge analysed data can be classified as new features, such as AI model update.

9.2.2 Procedures for control TE

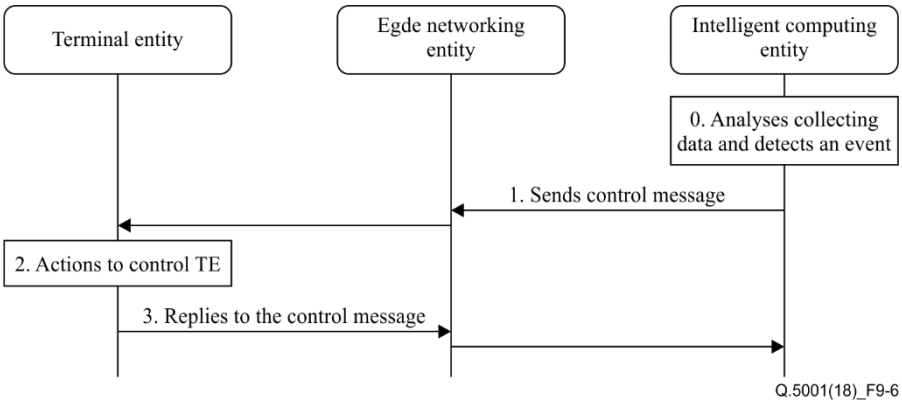


Figure 9-6 – Procedures for control TE

0. According to edge analysis for collecting data, abnormal situation or some event has been predicted.
1. ICE sends a request message to command an action.

   NOTE 1 – Types of actions can be defined by service profile.

   NOTE 2 – For video surveillance systems, ICE can control a surveillance camera directly.
2. TE takes the action immediately.

   NOTE 3 – Similar to Note 2, the camera can encode high-quality video in prediction time directly.
3. TE sends a reply to this request to ICE.

9.2.3 Procedures for edge analytic model update

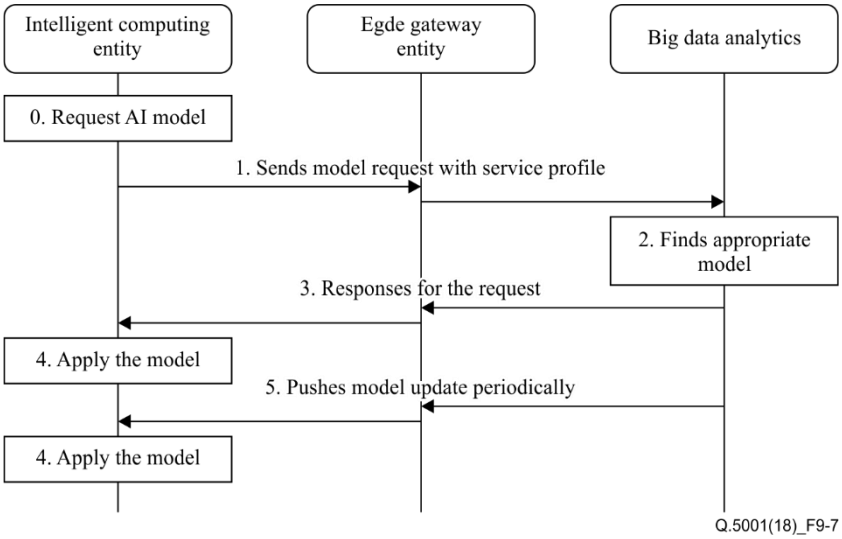


Figure 9-7 – Procedures for edge analytic model update

0. Before achieving edge analytics, ICE is required to request an AI model from a big data analytics on cloud server or a model repository in advance.
1. ICE sends a request message for AI model with a service profile.

   NOTE 1 – A model serving procedure could be substituted for the request.
2. In line with the service profile, appropriate AI model is located at the big data analytics server or model repository.
3. The server sends request attaching the AI model as well as parameters.

NOTE 2 – This procedure is similar to the model serving operation. More detail of the serving operation could be out of scope.

4. The model is applied to ICE to proceed with edge analytics.
5. After a reasonable period of time, the big data analytics server can push to update model.

   NOTE 3 – In other cases, the ICE can explicitly request new AI model.
6. ICE applies the model.

9.3 Characteristics of signalling messages

This clause describes signalling messages to allow the signalling architecture to achieve edge networking and intelligent data processing.

9.3.1 Signalling messages for networking functions

To support edge networking, all message flows consist of the following three messages:

– **Connectivity request and response messages** enable TE or data to be accessible from ICEs. These messages are used in the same procedure as identity registration at TE. But ICE can request to connect other ICEs located in different IEC, so this request is contained in the lookup message through EME. To establish link connectivity between all of entities, low layer protocols should be exchanged, but these protocols specifications are outside the scope of this Recommendation.

– **Identity registration and lookup messages** are used to register TE's own identity or data identity to EME through ICE. To interwork with other IECs, EGE's identity as a representative IEC is stored in EME. Regarding lookup message, a data consumer wants to find an appropriate data producer's identity, so that the response message contains data producer's identity.

– **Data request and response messages** are proceeded by data identity. Usually the identity can be compatible with hierarchical URL or URI, thus the identity can be routable by itself. In IEC, an ICE aggregates raw data in advance as data depot to be provided to a data consumer.

9.3.2 Signalling messages for intelligent data processing

To support intelligent data processing, first the analytic model is required to analyse data gathered through data forwarding. Also, as a result of edge analytics, IEC sends a control message to the TE.

– **Data forwarding message** is basically to cooperate with the data request message, but it is possible to push data without specific data request.

– **Control message** is to control the TE's actions, thus TE's operations should be registered in advance.

– **AI model request and update messages** interwork with a big data server or model repository at cloud computing. The request message contains a service profile to find an appropriate AI model. Periodically, the model is updated through big data analysis.

9.3.3 Message format

All messages are used in the type-length-value (TLV) encoding format, thus it can be encoded in two-byte type and length fields. Each message consists of the message header and the payload. The message header format is described in Table 9-1.

Table 9-1 – Message header fields

Field name	Description
Version	This indicates version of message format
Message type	This uniquely specifies the type of message
Message length	This specifies the length of total message
Reserve field	This reserves option filed
Header length	This specifies the length of header
Identity	This uniquely identifies the name of entity or data Note – Identity is defined by TLV
Optional header	This reserves optional header Note – This is defined by TLV

The message type can be one of the following:

These message types are defined by networking functions:

– T_REQ_CONNECT
– T_REP_CONNECT
– T_REG_IDENTITY
– T_REP_REG_IDENTITY
– T_GET_IDENTITY
– T_REP_GET_IDENTITY
– T_REQ_DATA
– T_REP_DATA

These message types are defined by data processing:

– T_PUSH_DATA
– T_RES_PUSH_DATA
– T_REQ_CONTROL
– T_REP_CONTROL
– T_REQ_MODEL
– T_REP_MODEL
– T_UPDATE_MODEL
– T_REP_UPDATE_MODEL

As shown in Table 9-2, message payload contains payload type and payload data. The payload data can contain other payload fields by using a piggyback message type.

Table 9-2 – Message payload fields

Field name	Description
Payload type	This identifies the name of data.
Payload length	This specifies the length of payload.

Table 9-2 – Message payload fields

Field name	Description
Payload data	This is payload data. (Note)
NOTE – This payload data can contain other message payload fields.	

The payload type can be one of the following:

– P_DATA

– P_IDENTITY

– P_RESPONSE_CODE

– P_MESSAGE

In particular, the response for request messages has a response code to indicate the result. Table 9-3 lists response codes and their semantics.

Table 9-3 – Response code and semantics

Response code	Semantics
REP_OK	The request is accepted and was successful.
REP_FAIL	The request cannot be accepted.
REP_CONNECT_REFUSE	The connection request cannot be accepted.
REP_IDENTITY_AUTH	The identity request and lookup are accepted and authorized.
REP_CONTROL_REFUSE	The control request cannot be accepted.

10 Deployments

The IEC will support deployment scenarios where the IEC is deployed either at the long term evolution (LTE) base station evolved node base station (eNB) site, or at the wireless local area network (WLAN) access point site, or at a multi-technology (LTE/WLAN) cell aggregation site. The multi-technology (LTE/WLAN) cell aggregation site can be located indoor within an enterprise (e.g., hospital, factory), or indoor/outdoor for a special public coverage scenario (e.g., stadium, shopping mall, department store) to control a number of local multi-technology (LTE/WLAN) access points providing radio coverage to the premises. This deployment option enables the direct delivery of locally-relevant, fast services from LTE base station or WLAN access point clusters. Figure 10-1 depicts the deployment scenarios of an IEC.

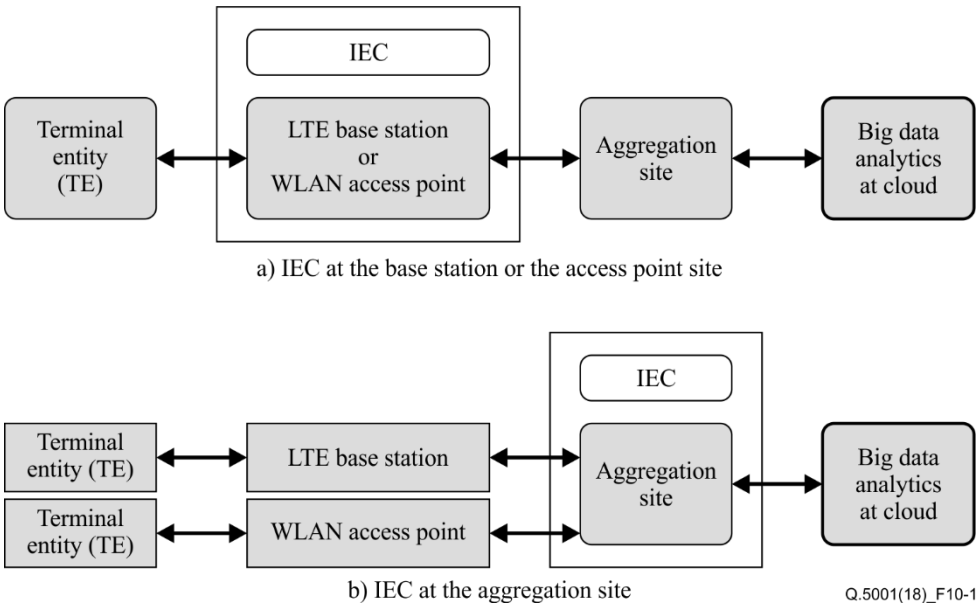


Figure 10-1 – Deployments for IEC

11 Security considerations

The intelligent edge computing shall provide a secure environment for running services for the following actors: the user, the network operator, the third-party application provider, the application developer, the content provider, and the platform vendor. The intelligent edge computing shall only provide an edge application with the information for which the application is authorized.

Appendix I

Use cases of intelligent edge computing

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

Autonomous network control is the network automation part to improve the performance of the network itself. Autonomous network control may consist of autonomous quality of experience (QoE) policy control, intelligent traffic classification and so on. Through network automation, it is possible to automatically process the portion of the existing network where the operator enters the policy directly to provide QoE. Ambient intelligence is the internalization part of the network intelligence for using ubiquitously the function and improving the performance of machine learning. Ambient intelligence consist of selective data collection, dynamic storage, real-time trust data pre-process and so on. In particular, edge node can receive IoT data at the near source devices. Therefore, this concept can increase data utilization by sending data reliably and promptly to facilitate data collecting, loading and processing functions to be performed in each IEC [b-ETSI GS MEC 004].

I.1 Mobile video delivery optimization using throughput guidance for transmission control protocol (TCP)

Media delivery is nowadays usually done via hypertext transfer protocol (HTTP) streaming which in turn is based on the transmission control protocol (TCP). The behaviour of TCP, which assumes that network congestion is the primary cause for packet loss and high delay, can lead to the inefficient use of a cellular network's resources and degrade application performance and user experience. The root cause for this inefficiency lies in the fact that TCP has difficulty adapting to rapidly varying network conditions. In cellular networks, the bandwidth available for a TCP flow can vary by an order of magnitude within a few seconds due to changes in the underlying radio channel conditions, caused by the movement of devices, as well as changes in system load when other devices enter and leave the network.

In this use case, an edge application for network analytics provides a suitably equipped backend video server with a near real-time indication on the throughput estimated to be available at the radio downlink interface in the next time instant. The video server can use this information to assist TCP congestion control decisions. With this additional information, TCP does not need to overload the network when probing for available resources, nor does it need to rely on heuristics to reduce its sending rate after a congestion episode.

The throughput guidance is an application-specific figure, which gives the video server a hint about the bitrate that can be expected to be available for its use during the upcoming time interval.

I.2 Active device location tracking

This use case enables real-time, network measurement based tracking of active global positioning system (GPS) (independent and network determined) terminal equipment using geo-location algorithms.

This provides an efficient and scalable solution with local measurement processing and event based triggers. It enables location based services for enterprises and consumers (e.g., on opt-in basis), for example, in venues, retail locations and traditional coverage areas where GPS coverage is not available. Services can include mobile advertising, 'Smart City', footfall analysis, campus management, crowd management, personal management, etc.

I.3 Bandwidth allocation manager for applications

In some cases, different sessions running in parallel in the same application can each have specific bandwidth requirements. As all these applications and application sessions are competing over the

same shared bandwidth resources, it is suggested that a central bandwidth resource allocator exists on the edge node. The proposed function can include the following:

– an application programmable interface (API) enabling all registered application to statically and/or dynamically register for specific bandwidth allocation;
– an interface with the network information service to receive network conditions and available bandwidth;
– the capability to calculate optimal bandwidth allocation per session/application according to available and required bandwidths;
– the capability to manage the bandwidth allocated to each of the sessions/applications according to the calculations.

I.4 Video caching, compression and analytics service chaining

Consider the use case where traffic that is sent from the content caching application to a TE is steered first through the video compression application and then through the video analytics application. When the uplink request arrives at the edge node, the content request is routed to the content caching application in order to retrieve the content.

Once the content is identified, the user traffic needs to be passed to the video compression and video analytics application before it can be delivered to the end user.

The edge node needs to support this scenario, whereby it will classify the traffic and then steer the traffic through multiple applications.

I.5 Application computation off-loading

In the application computation off-loading use-case, the edge node executes compute-intensive functionalities with high performance instead of TE. By providing rich computation resources on an edge node, application computation can be off-loaded to the edge node to be accelerated even if a user uses relatively low performance devices, and user experience can be satisfied regardless of the type of TE.

This use-case is effectively used for especially computation-hungry applications such as graphical rendering (high-speed browser, artificial reality, 3D game, etc.), intermediate data-processing (sensor data cleansing, video analysing, etc.), and value-added services (translation, log analytics, etc.). An example of application computation offloading is the edge accelerated browser (EAB). Most parts of the browsing functions, such as web contents evaluation, rendering and optimized transmission, are off-loaded to the edge node, while the TE just renders reconstituted browser graphics on its display. This can transfer a compute-intensive process from a TE to an edge node to accelerate an application and make rich applications available on various types of devices.

I.6 Data path off-loading

In an IoT environment, lots of sensing data will be generated and large volumes of data can be forwarded to an analysis server in the cloud. Thus, edge devices including access point, switch, actuator, mobile phone, etc., have been faced on overhead for data transmission. Typically, mobile crowd sensing (MCS) is a sort of crowdsourcing application involving many smart phones in different locations. However, in terms of sensing data transmission, the MCS can provide data off-loading where data are forwarded to an aggregation server through various paths.

In this use-case, a smart device with significant computation, communication and sensing capabilities can be a kind of intelligent edge node. For instance, a sort of temperature data will be generated to a single digit number, but the MCS application can collect the single unit data and it will process and analyse collecting data to generate more meaningful information, such as a graph of temperature changes per hour or temperature changes with time and place. Additionally, the MCS application can

choose a close-by aggregation server or one in the cloud. Thus, the aggregation server located in an edge network can process and analyse collecting data from MCS application by using an intelligent analysis platform such as a machine learning platform.

I.7 Intelligent data processing and filtering

This use case groups a number of innovative services for the operator or third-party vendors based on the gathering of large amounts of data (video, sensor information, etc.) from devices analysed through a certain amount of processing to extract meaningful information before being sent towards central servers. Applications might run in a single location, or be spread over a given area (e.g., campus coverage) or even in the whole network. In order to support the constraints of the operator or the third party requesting the service, the applications might have to be run on all requested locations.

This use case describes an application running on an edge node that receives a very large amount of information from devices and sensors connected to the edge node. The application then processes the information and extracts the valuable metadata, which it sends to a central server. A subset of the data might be stored locally for a certain period for later verification.

A number of service scenarios can be enabled such as monitoring of an area for specific events, abandoned luggage, authorized access (e.g., with face recognition), car park car monitoring, massive sensor data pre-processing, smart city, etc.

I.8 Smart construction monitoring system using machine learning techniques

There are many dangerous elements in construction sites, such as noise, gas leaks and vibrations for which alerts are triggered. Real-time monitoring systems can detect the alerts using machine learning techniques (DL, RL) that can provide more effective solutions and approaches to recognize dangerous construction elements.

To monitor these elements in construction sites, there should be closed circuit television (CCTV) systems operating locally, and continuously broadcasting. It is usually ineffective and a waste of resources to transmit still images even if the CCTV constantly broadcasts them in high definition. However, if an alert is detected due to dangerous elements, the streaming should be converted to high quality streaming data to rapidly show and identify the dangerous situation. From a technical approach, deep learning (DL) is one of the solutions to automatically detect these kinds of dangerous situations in advance using prediction. It can provide the data, including the high-rate streaming video, to quickly control other risks. Reinforcement learning (RL) is additionally an important approach to efficiently manage and monitor the given dataset in real time.

Appendix II

Related works of intelligent edge computing

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

Basic edge node technology in various shapes and forms is available on the market from several telecom vendors since 2013. Currently, all vendors work on edge solutions, and many telecom operators are conducting trials with edge technology. Some operators have already launched edge services and the main use cases are edge caching services. There are a number of edge computing initiatives in several industries. ETSI Industry Specification Group "Mobile Edge Computing (MEC)" actively develops requirement, architectures and specifications [b-ETSI GS MEC 002] [b-ETSI GS MEC 003] [b-ETSI GS MEC 004]. Examples of IT industry initiatives are OpenFog [b-OpenFog], OpenStack [b-OpenStack], etc.

The following gaps currently exist:

– There is a need for an open approach regarding the positioning of the edge component (base station, core net, elsewhere).

– To guarantee the success of IEC, it is important to work closely with the information technology (IT) industry.

– Active engagement should be pursued with several vertical industries that will use edge computing in order to provide an intelligent edge API and reference platform for intelligent edge computing, and align it with the IT industry and the telecoms industry.

Table II.1 – Related works of intelligent edge computing

Item	MEC	Fog	IEC
Allows multi-tenancy	O	O	O
Can be physically co-located	O	O (Access points, base stations, traffic aggregation points, routers, switches)	O
Extends Cloud	Δ (Can be, but does not need to be)	O	Δ (Can be, but does not need to be)
Focus on on-line analytics	X	O	O
Inspired by	Low Latency and Device Context	IoT	Massive data forwarding and analysing, AIaaS
Located between end device and core network	O	O	O (plus can run on end device)
N-tier hierarchy	N – 2 or 3	N=3	N – 3 or 4
Uses Virtual IaaS	O	O	O
Used with	Wireless Access	Wireless Access	Fixed and Wireless Access

Bibliography

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

[b-ETSI GS MEC 002] ETSI GS MEC 002 V1.1.1 (2016), *Mobile Edge Computing (MEC); Technical Requirements.*

[b-ETSI GS MEC 003] ETSI GS MEC 003 V1.1.1 (2016), *Mobile Edge Computing (MEC); Framework and Reference Architecture.*

[b-ETSI GS MEC 004] ETSI GS MEC-IEG 004 V1.1.1 (2015), *Mobile Edge Computing (MEC); Service Scenarios.*

[b-OpenFog] Open Fog Consortium
https://www.openfogconsortium.org/

[b-OpenStack] Open Stack Foundation
https://www.openstack.org/

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

Printed in Switzerland
Geneva, 2018