5G ToB Service Experience Standard White Paper

Jan 2021

Contents

1 Overview....4

2 5G ToB Service Introduction....5

2.1 Service Development.... 5
2.2 Reference Protocols.... 6
2.2.1 3GPP TR38.913.... 6
2.2.2 ITU-R IMT-2020.... 8
2.2.3 ETSI TR 103 702.... 10
2.2.4 5G-PPP....11
2.3 Service Introduction.... 17
2.3.1 Video Transmission.... 17
2.3.2 Industrial Park.... 26
2.3.3 Industrial Automation.... 28
2.3.4 Industrial UAV.... 32
2.3.5 FWA Service.... 33
2.3.6 Smart City.... 35
2.3.7 Massive Connectivity Services.... 37
2.4 Related Technologies.... 39
2.4.1 UDP.... 39
2.4.2 RTSP.... 43
2.4.3 IoT Protocols.... 46

3 5G ToB Service Characteristic Analysis....51

3.1 Multi-Media Transmission.... 51
3.1.1 HD Live Broadcast at Site C.... 51
3.1.2 HD Live Broadcast at Site K.... 56
3.1.3 Video Surveillance at Site X.... 59
3.2 Interactive Service Behavior.... 62
3.2.1 PLC-PNIO.... 62
3.2.2 PLC-S7Comm.... 64
3.3 FWA Service.... 70

4 5G ToB Service Modeling Framework....74

4.1 ToB & B2C Modeling Differences.... 74
4.2 ToB Modeling Method Exploration.... 75
4.2.1 Fine-grained Spatio-temporal Modeling.... 75
4.2.2 Scenario-based Event-driven Modeling.... 77
4.3 ToB Modeling Frame.... 82
4.3.1 Indicator-driven Modeling Framework.... 82
4.3.2 Event-driven Modeling Framework.... 83

5 5G ToB Service Indicator System....84
5.1 Uplink Multimedia Transmission Service 84
5.1.1 Impact Factor.... 84
5.1.2 Indicator System.... 86
5.1.3 Modeling Method.... 88
5.1.4 Experience Baseline.... 92
5.2 Downlink Multimedia Transmission Service.... 100
5.2.1 Impact Factor.... 100
5.2.2 Indicator System.... 101
5.2.3 Modeling Method.... 101
5.2.4 Experience Baseline.... 101
5.3 AR Service.... 101
5.3.1 Impact Factor.... 101
5.3.2 Indicator System.... 104
5.3.3 Modeling Method.... 105
5.3.4 Experience Baseline.... 106
5.4 Real-Time Interaction Service.... 106
5.4.1 Impact Factor.... 106
5.4.2 Indicator System.... 107
5.4.3 Modeling Method.... 107
5.4.4 Experience Baseline.... 108
5.5 Massive Connectivity Service 110
5.5.1 Impact Factor....110
5.5.2 Indicator System....111
5.5.3 Modeling Method....112
5.5.4 Experience Baseline....113
5.6 FWA Service....113
5.6.1 Impact Factor....113
5.6.2 Indicator System....114
5.6.3 Modeling Method....116
5.6.4 Experience Baseline....116
6 References 119
Abbreviations and Acronyms 120

1 Overview

With 5G still at the outset, foreseeable service scenarios including cloud virtual reality (VR) and augmented reality (AR), high-definition (HD) video, livelihood, and industrial campus are currently in demand. Many more scevice scenarios are yet to be extensively used such as Massive Machine-Type Communications (mMTC) and ultra-reliable low-latency communication (URLLC). This document analyzes the service characteristics, network requirements, metric systems, and modeling algorithms based on foreseeable demands at the initial phases of 5G construction and promotion.

2 5G ToB Service Introduction

2.1 Service Development

The International Telecommunication Union - Radio communication Sector (ITU-R) has defined three major service scenarios of 5G, as illustrated in Figure 2-1.

Figure 2-1 Three major service scenarios of 5G

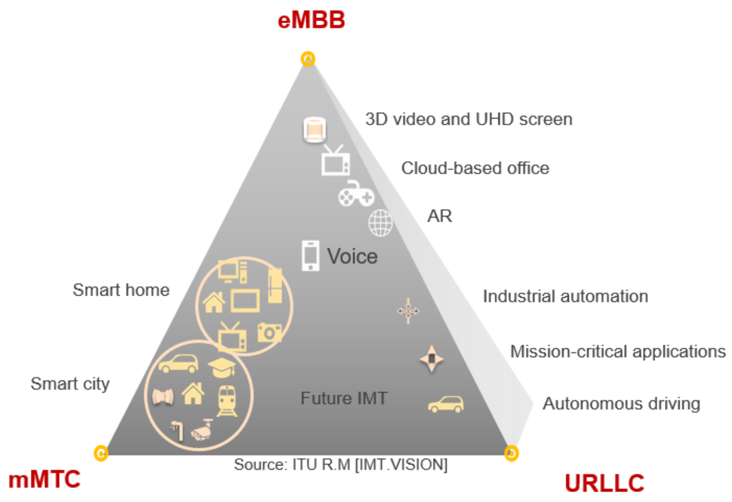
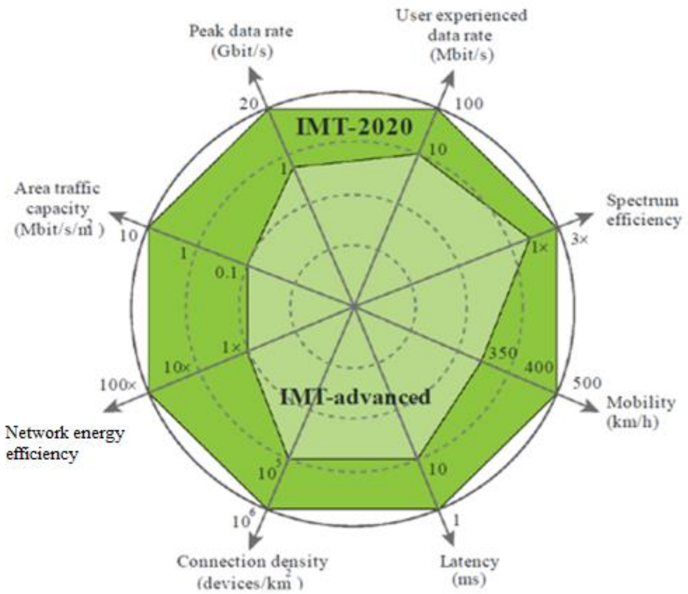


Figure 2-2 illustrates the key capabilities of 5G.

Figure 2-2 Key capabilities of 5G



5G business-to-business (ToB) development falls into three stages:

- Early stage: Enhanced Mobile Broadband (eMBB) dominates, partially with ultra-reliable low-latency communication (URLLC) services (not necessarily within 10 ms). Massive Machine-Type Communications (mMTC) applications are basically none in existing 5G projects.
- Development stage: URLLC applications gradually increase, and eMBB applications further develop.
- Mature stage: mMTC applications begin to rise and grow more complex. Behaviors are intertwined all-round, and everything is connected.

ToB buyers aspire a business-to-consumer (B2C) experience in view that suppliers offer more personalized services.

ToB development now faces the following challenges:

- Enterprise private network solutions are insufficiently standardized.
- There are various protocols for industrial applications.
- Passive measurement of network performance poses numerous difficulties.

The majority of operators lack capabilities for end-to-end (E2E) solution design and delivery. Based on ToB service scenarios, this document maps service scenarios onto typical transmission service behaviors, and builds models for each service patterns. Metric systems, modeling methods, and theoretical basics are also provided to help design solutions for evaluating, monitoring, and optimizing the ToB service experience.

2.2 Reference Protocols

This section describes the supporting protocols and specifications for evaluating 5G ToB service experience.

2.2.1 3GPP TR38.913

Some of the key performance indicators (KPIs) of 5G network services are listed below:

Definition	Description
Peak data rate	Indicates the highest theoretical data rate, which is the received data bits assuming error-free conditions assignable to a single mobile station, when all assignable radio resources for the corresponding link direction are utilized.
Peak spectral efficiency	Indicates the highest theoretical data rate (normalized by bandwidth), which is the received data bits assuming error-free conditions assignable to a single mobile station, when all assignable radio resources for the corresponding link direction are utilized.
Bandwidth	Indicates the maximal aggregated total system bandwidth.
Control plane latency	Indicates the time taken to move from a battery efficient state (such as idle) to the start of continuous data transmission state (such as active). The target for control plane latency should be 10 ms.

Definition	Description
User plane latency	Indicates the time taken to successfully deliver an application layer packet or message from a service data unit (SDU) ingress point on Layer 2 or 3 to an SDU egress point on Layer 2 or 3 over the uplink and downlink radio interfaces. The target for user plane latency should be 0.5 ms for URLLC and 4 ms for eMBB, for both uplink and downlink.
Latency for infrequent small packets	Indicates the time taken to successfully deliver a packet or message from an SDU ingress point on Layer 2 or 3 of a mobile device to an SDU egress point on Layer 2 or 3 in the radio access network (RAN), when the mobile device starts from its most "battery efficient" state. This KPI measures infrequent transmission of small packets or messages over the application layer.
Mobility interruption time	Indicates the shortest time duration supported by a system during which a user terminal cannot exchange user-plane packets with any base station during transitions. The target for mobility interruption time should be 0 ms.
Inter-system mobility	Indicates the ability to support mobility between the International Mobile Telecommunications 2000 (IMT-2000) system and at least one IMT system.
Reliability	Reliability can be evaluated by the success probability of transmitting *X* bytes within a certain delay, which is the time taken to deliver a small data packet from an SDU ingress point on Layer 2 or 3 to an SDU egress point on Layer 2 or 3 over the radio interface, at a certain channel quality (such as coverage-edge).
Coverage	Indicates the maximum coupling loss (MCL) in uplink and downlink between a device and a base station (antenna connectors for a data rate of 160 bps, where the data rate is observed at the egress or ingress point of the radio protocol stack in uplink and downlink). The target for coverage should be 164 dB.
Extreme coverage	The coupling loss is the total long-term channel loss over the link between a UE's antenna ports and an eNodeB's antenna ports, and includes in practice antenna gains, path loss, shadowing, body loss, and others.
UE battery life	Indicates a UE's battery life without recharge. For mMTC, the UE battery life in extreme coverage should be based on the activity of mobile originated data transmission consisting of 200 bytes uplink per day followed by 20 bytes downlink from a MCL of 164 dB, assuming a stored energy capacity of 5 Wh.
UE energy efficiency	Indicates the capability of a UE to sustain much better mobile broadband (MBB) data rates while minimizing the UE modem energy consumption.

Definition	Description
Spectral efficiency per cell or transmission and reception point (TRxP)	TRxP spectral efficiency is the aggregate throughput of all users (the number of correctly received bits, namely the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time) divided by the channel bandwidth divided by the number of TRxPs.
Area traffic capacity	Indicates the total traffic throughput served per geographic area (in Mbps per m^2). This KPI can be evaluated using the full buffer or non-full buffer model. Area traffic capacity (bps/m^2) = Site density (site/m^2) x Bandwidth (Hz) x Spectral efficiency (bps/Hz/site)
User experienced data rate	The user experienced data rate is 5% of the user throughput, for non-full buffer traffic. User throughput (during active time) is the size of a burst divided by the time between the arrival of the first packet of a burst and the reception of the last packet of the burst. User experienced data rate = 5% user spectral efficiency x Bandwidth
5% user spectral efficiency	Indicates the 5% point of the cumulative distribution function (CDF) of the normalized user throughput. The (normalized) user throughput is the average user throughput (the number of correctly received bits by users).
Connection density	Indicates the total number of devices fulfilling a target quality of service (QoS) per unit area (per km^2), where the target QoS is to ensure a system packet loss rate less than 1% under a given packet arrival rate and packet size. The packet loss rate is obtained by the following formula: Number of packets in outage/Number of generated packets, where a packet is in outage if it failed to be successfully received by the destination receiver beyond a packet dropping timer.
Mobility	Indicates the maximum user speed at which a defined QoS can be achieved (in km/h).
Network energy efficiency	Indicates the capability to minimize the RAN energy consumption while providing an improved area traffic capacity.

2.2.2 ITU-R IMT-2020

The standard protocols of 3G and 4G were developed by regional standards organizations such as the 3rd Generation Partnership Project (3GPP). The ITU's influence on 3G, 4G, and 5G standards, however, predominantly lies in proposing market demands, constructing blueprints and visions, and creating global consensus and ecosystem. The ITU-R has just announced the performance requirements for 5G, or International Mobile Telecommunications 2020 (IMT-2020), consisting of:

1. Peak data rate per cell
   - Downlink: 20 Gbps
   - Uplink: 10 Gbps
2. Peak spectral efficiency per cell
   - Downlink: 30 bps/Hz

- Uplink: 15 bps/Hz

3. Experienced data rate per user
   - Downlink: 100 Mbps
   - Uplink: 50 Mbps
4. 5% spectral efficiency

Test Environment	Downlink (bps/Hz)	Uplink (bps/Hz)
Indoor hotspot – eMBB	0.3	0.21
Dense urban – eMBB*	0.225	0.15
Rural – eMBB	0.12	0.045
*: This requirement will be evaluated under the macro TRxP layer of the dense urban – eMBB test environment as described in Report ITU-R M.[IMT-2020.EVAL].		

5. Average spectral efficiency

Test Environment	Downlink (bps/Hz/TRxP)	Uplink (bps/Hz/TRxP)
Indoor hotspot – eMBB	9	6.75
Dense urban – eMBB*	7.8	5.4
Rural – eMBB	3.3	1.6
*: This requirement applies to the macro TRxP layer of the dense urban – eMBB test environment as described in Report ITU-R M.[IMT-2020.EVAL].		

6. Data throughput per unit area
   - Downlink and indoor hotspot: 10 $Mbps/m^2$
7. User plane latency
   - eMBB: 4 ms
   - URLLC: 1 ms
8. Control plane latency: 20 ms
9. Connection density: 1 million devices per km^2
10. Network energy efficiency
    - Effective data transmission under loads
    - Low energy consumption without data transmission
11. Reliability: $1–10^{-5}$
12. Mobility

Mobility levels:

Indoor Hotspot – eMBB	Dense Urban – eMBB	Rural – eMBB
Stationary, pedestrian	Stationary, pedestrian, vehicular (up to 30 km/h)	Pedestrian, vehicular, high speed vehicular

Data rates (normalized by bandwidth) over traffic channel links:

Test Environment	Normalized Traffic Channel Link Data Rate (bps/Hz)	Mobility (km/h)
Indoor hotspot – eMBB	1.5	10
Dense urban – eMBB	1.12	30
Rural – eMBB	0.8	120
	0.45	500

13. Mobile interruption time (MIT): The MIT includes the time required to perform any RAN process applicable to the candidate radio interface technology (RIT) or a set of RITs (SRIT), radio resource control signaling protocol, or other message exchanges between a mobile station and the RAN. The minimum MIT should be 0 ms.
14. System bandwidth: It should be at least 100 MHz. An RIT or SRIT should support bandwidths up to 1 GHz in order to operate in higher frequency bands (for example, above 6 GHz).

2.2.3 ETSI TR 103 702

In terms of 5G service metric system specifications, Huawei has proposed to the ETSI standards for the VR experience metric system. Note these are specifications yet to be officially released.

Service Type	Service Indicator	Indicator Requirement
Terminal	Terminal resolution	2K–4K
Strong-interaction cloud VR services	Content resolution (equivalent full-view resolution)	2K–4K (equivalent full-view: 4K–8K)
	Color depth (bits)	8
	Coding mode	H.264, H.265
	Bitrate (Mbps)	≥ 40
	Frame rate (FPS)	50–90
	Field of view (FoV) (degrees)	90–110
	Interactive latency (ms)	≤ 100
	MTP (ms)	≤ 20
	Valid frame rate	100%
Cloud VR video	Content full-view resolution	4K–8K

Service Type	Service Indicator	Indicator Requirement
services	Color depth (bits)	8
	Coding mode	H.264, H.265
	Bitrate (Mbps)	≥ 40
	Frame rate (FPS)	30
	FoV (degrees)	90–110
	Interactive latency (ms)	≤ 100
	Initial buffer latency (s)	≤ 1
	Stalling duration ratio	0
	Pixelization duration ratio	0

2.2.4 5G-PPP

5G Infrastructure Public Private Partnership (5G PPP) is a joint initiative by the European Commission and the European ICT industry (including ICT manufacturers, telecom operators, service providers, SMEs and research institutions).

5G Mobile Network Architecture (5G-MoNArch) for diverse services, use cases, and applications in 5G and beyond is a project initiated by the 5G PPP.

5G-MoNArch brings forward the next-step development of the 5G mobile network architecture. It fully integrates network functions required by industries, media and entertainment, and smart city into the overall architecture, so that the mobile network architecture can be used for practical applications. To verify the feasibility and applicability of concepts developed by it in real environments, 5G-MoNArch is built on two means: project testing platform and verification framework.

- Testing platform:

  5G-MoNArch has implemented two testing platforms: the Smart Sea Port platform in Hamburg Germany and the Touristic City platform in Turin Italy. Both platforms have helped promote the verification of performance targets. They are now benchmarks for technical and economic feasibility verification.
- Verification and confirmation: In order to quantify the technical and socio-economic benefits of technologies developed by it, 5G-MoNArch has defined a framework, which includes a process and a set of technical, commercial, and economic KPIs. The framework has been evaluated according to three defined cases.

5G network service KPIs:

Definition	Description
General KPI	

Definition	Description
Area traffic capacity (based on 3GPP/ITU-R)	The total traffic throughput served per geographic area (in bps/m^2). This KPI can be evaluated by two different traffic models: • By a full buffer model: The calculation of the total traffic throughput served per geographic area is based on full buffer traffic. • By a non-full buffer model: The total traffic throughput served per geographic area is calculated. However, the user experienced data rate needs to be evaluated at the same time using the same traffic model in addition to the area traffic capacity. The area traffic capacity is a measure of traffic volume a network can carry per unit area. It depends on site density, bandwidth, and spectral efficiency. In the case of full buffer traffic and a single-layer single-band system, it can be expressed as: Area traffic capacity (bps/m^2) = Site density (site/m^2) x Bandwidth (Hz) x Spectral efficiency (bps/Hz/site)
Availability (based on 3GPP/5G PPP/NGMN/ETSI)	Percentage value (%) of the amount of time a system can deliver services divided by the amount of time it is expected to deliver services in a specific area. The availability may be specific for a communication service. In this case, it refers to the percentage value of the amount of time the end-to-end (E2E) communication service is delivered according to an agreed QoS, divided by the amount of time the system is expected to deliver the E2E service according to the specification in a specific area. **NOTE** • The end point in "E2E" is assumed to be the communication service interface. • The communication service is considered unavailable if it does not meet the pertinent QoS requirements.
Bandwidth (based on 3GPP)	Indicates the maximal aggregated total system bandwidth.
Cell-edge user throughput (based on 3GPP)	Indicates the fifth percentile point of the CDF of user's average packet call throughput.
Connection density (based on 3GPP/ITU-R)	The total number of connected and/or accessible devices per unit area (per km^2). Connectivity or accessibility refers to devices fulfilling a target QoS, where the target QoS is to ensure a system packet loss rate less than [x]% under given packet arrival rate [l] and packet size [S]. The packet loss rate is equal to the number of packets in outage divided by the number of generated packets. A packet is in outage if this packet fails to be successfully received by the destination receiver beyond a packet dropping timer.
Coverage (based on 3GPP)	Indicates the MCL in uplink and downlink between a UE and a TRxP (antenna connectors for a data rate of [x] bps. The data rate is observed at the egress or ingress point of the radio protocol stack in each direction.

Definition	Description
Coverage area probability (based on 5G PPP)	Indicates the percentage of the area under consideration, in which a service is provided by a mobile radio network to an end user in a quality (such as the data rate, latency, or packet loss rate) that is sufficient for the intended application (QoS or QoE). The RAN may consist of a single radio cell or multiple cells. For services of different types and quality of service (QoS) or quality of experience (QoE) levels, the coverage area probability will also vary.
End-to-end latency (based on 3GPP/5G PPP)	Indicates the time to transfer a given piece of information from a source to a destination, which is measured at the communication interface, from the moment it is transmitted by the source to the moment it is successfully received at the destination. It is also referred to as one-trip-time (OTT) latency. Another latency measure is the round-trip-time (RTT) latency which refers to the time from when a data packet is sent from the transmitting end until acknowledgements are received from the receiving entity.
Energy efficiency (based on 3GPP/ITU-R)	It means to sustain a certain data rate while minimizing the energy consumption.
Latency for infrequent small packets (based on 3GPP)	Indicates the time to successfully deliver a packet or message from an SDU ingress point on Layer 2 or 3 at a UE to an SDU egress point on Layer 2 or 3 in the RAN, when the UE starts from its most "battery efficient" state. This KPI is a measure of infrequent transmission of small packets or messages over the application layer.
Mean time between failures (MTBF) (by ETSI)	Indicates the statistic mean uptime of a system or component before it fails.
Mean time to repair (MTTR) (by ETSI)	Indicates the statistic mean downtime before a system or component is back in operation again.
Mobility (based on 3GPP/ITU-R)	Indicates the maximum speed at which a defined QoS and seamless transmission between TRxPs which may belong to different deployment layers (namely multi-layer) and/or radio access technologies (namely multi-RAT) can be achieved (in km/h).
Mobility interruption time based on (3GPP/5G PPP)	Indicates the shortest time duration supported by a system during which a UE cannot exchange user-plane packets with any TRxP during transitions. This KPI is for both intra- and inter-frequency mobility as well as for mobility inside an air interface variant (AIV) or across AIVs.
Peak data rate (based on 3GPP/ITU-R/5G PPP)	Indicates the highest theoretical single-user data rate (in bps), assuming ideal, error-free transmission conditions, when all available radio resources for the corresponding link direction are utilized (excluding radio resources that are used for physical layer synchronization, reference signals or pilots, guard bands and guard times).

Definition	Description
Peak spectral efficiency (based on 3GPP)	Indicates the peak data rate normalized by the bandwidth applied. Higher frequency bands could have higher bandwidth but lower spectral efficiency, and lower frequency bands could have lower bandwidth but higher spectral efficiency. Thus, peak data rates cannot be directly derived from peak spectral efficiency and bandwidth multiplication.
Reliability (based on 3GPP/ITU-R/5G PPP/NGMN)	Indicates the percentage (%) of the amount of sent network layer packets successfully delivered to a given system node (including a UE) within the time constraint required by the targeted service, divided by the total number of sent network layer packets.
Resilience (based on ITU-R)	Indicates the ability of a network to continue operating correctly during and after a natural or man-made disturbance, such as the loss of mains power.
Service continuity (based on 3GPP)	Indicates the uninterrupted user experience of a service that is using an active communication when a UE undergoes an access change without the user noticing the change.
Spectral efficiency per cell or TRxP (based on 3GPP/ITU-R)	TRxP spectral efficiency indicates the aggregate throughput of all users (the number of correctly received bits, specifically the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time) within a radio coverage area (site) divided by the channel bandwidth divided by the number of TRxPs. A 3-sector site consists of 3 TRxPs. In the case of multiple discontinuous "carriers" (one carrier refers to a continuous chunk of spectrum), this KPI should be calculated per carrier. In this case, the aggregate throughput, channel bandwidth, and the number of TRxPs on the specific carrier are employed.
Spectrum and bandwidth flexibility (based on ITU-R)	Indicates the flexibility of the 5G system design to handle different scenarios, and in particular the capability to operate at different frequency ranges, including higher frequencies and wider channel bandwidths than today.
UE battery life (based on 3GPP)	Indicates the life time of the UE battery to be evaluated without recharge. Note: For mMTC, 3GPP proposed that the UE battery life in extreme coverage shall be based on the activity of mobile originated data transmission consisting of 200 bytes uplink per day followed by 20 bytes downlink from MCL of 164 dB, assuming a stored energy capacity of 5 Wh.

Definition	Description
User experienced data rate (based on 3GPP/ITU-R)	It can be evaluated for non-full buffer traffic and for full buffer traffic. However, non-full buffer system level simulations are preferred for the evaluation of this KPI responsible of respective deployment scenarios and using burst traffic models. For non-full buffer traffic, the user experienced data rate is 5% of the user throughput. User throughput (during active time) is the size of a data burst divided by the time between the arrival of the first packet of a burst and the reception of the last packet of the burst. For full buffer traffic, the user experienced data rate is calculated as: User experienced data rate = 5% user spectral efficiency x Bandwidth
User plane latency (based on 3GPP/5G PPP)	Indicates the time to successfully deliver an application layer packet or message from an SDU ingress point on Layer 2 or 3 to an SDU egress point on Layer 2 or 3 over the uplink and downlink radio interfaces.
Resource Elasticity KPIs	
Availability	Indicates the relative amount of time that the function under study produces the output that it would have produced under ideal conditions, with a specific focus on the resource provisioning.
Cost efficiency gain	It measures the average cost of deploying and operating the network infrastructure to support the foreseen services. An elastic system should be able to be optimally dimensioned such that fewer resources are required to support the same services. Additionally, in lightly loaded scenarios, the elastic system should avoid using unnecessary resources and reduce the energy consumption.
Elasticity orchestration overhead	Indicates the amount of resources required for realizing orchestration functions, namely functions that enable network function (NF) elasticity, including the re-placement of a virtual network function (VNF), and are not part of the traditional architecture. An example could be the vector that includes the amount of CPU, RAM, and the amount of networking resources consumed by the orchestration function.
Minimum footprint	Given a set of resources to execute a function, the minimum footprint indicates the set of combinations of these resources that are needed to produce any output. Depending on the heterogeneity of these resources, it may be the case that there is a "region" of minimum footprints, which includes all the possible combinations of resources that results in a successful execution of the function.
Multiplexing gains	Indicates the number and kind of functions that can run in parallel over the same set of resources with a certain performance level.
Performance degradation function	This KPI characterizes the relation between a reduction in the available resources (from 100% until the minimum required) and the reduction in performance of a function. In this case, an elastic NF should achieve graceful performance degradation, avoiding abrupt breakdown under peaks.

Definition	Description
Rescuability	When a resource shortage occurs, scaling out or up the virtual machines (VMs) that are executing VNFs is the most likely solution to be adopted. Still, re-orchestration processes usually operate at larger time scales.
Resource consumption	Given a resource (CPU, RAM, or others), its consumption indicates the percentage of time it is occupied because of the execution of a function.
Resource savings	Indicates the amount and type of resources consumed by an elastic function to perform a successful operation as compared to its inelastic counterpart (for example, the percentage of saved resources while providing 99% of the performance of the inelastic counterpart).
Response time	Indicates the time required for resources to be provisioned when demand changes. The shorter the response time is, the greater the elasticity.
Resource utilization efficiency	It is a way to measure how resources are efficiently utilized to provide the desired output. An elastic system should be able to lead to a larger resource utilization efficiency, since it can deploy a higher number of VNFs over the same physical infrastructure.
Service creation time	Indicates the time from the arrival of a request to set up a network slice at the network operator's management system until the slice is fully operational.
Time for reallocation of a device to another slice	Indicates the duration from the request to connect a terminal device to a certain network slice until this device can start communication.
Application-specific KPIs	
Frame rate judder	$$\frac{n/75}{\sum_1^n t_n \epsilon (t > \frac{1}{75})}$$ where *t* is the time required for each frame to render and *n* the total number of rendered frames. The formula represents the percentage of time during a virtual reality (VR) application where the framerate is less than 75 frames per second. Minimizing this time reduces the probability of motion sickness.
Maximum number of simultaneously active IoT devices	It is expected that in the future, cargo containers will be equipped with smart sensors monitoring and reporting environmental conditions (such as the temperature, humidity, and bumps) online during their journey. A container ship today can carry up to 20,000 containers. When such a ship coming from overseas enters the coverage area of the very first mobile radio cell, possibly all 20,000 containers will attempt to access the radio cell almost simultaneously. This KPI measures the maximum number of sensors within a given deployment area that can be supported by a network slice.
Task success rate	Indicates the percentage of correctly completed tasks by users.

Definition	Description
Time on task	Indicates the task completion time or task time. This KPI is basically the amount of time it takes for a user to complete a task, expressed in minutes and seconds.
Use of search vs. navigation	This is a valuable metric for evaluating the efficiency of information architecture and navigation. Usually when users try to find something through navigation and get lost, search is their final option. Using this KPI, the user perception of network failures can be measured and then correlated to the underlying problem.

2.3 Service Introduction

The use cases for the next generation communications have expectedly higher requirements on QoS. This is in terms of a higher data rate and larger network throughput, for eMBB, URLLC, and mMTC.

The key requirements are as follows:

- 5G networks offer multiple service-specific QoSs, compared to only one QoS for the entire network.
- The network slice management and orchestration (MANO) layers use QoS to manage current network slice performance. It additionally allocates necessary resources in the virtual environment to different VNFs in different domains (RAN, core network and transport network).
- Effective E2E QoS negotiation requires application and service awareness at multiple points on various networks.
- Machine learning and artificial intelligence (AI) are key to enabling multi-point data sources and real-time flow analysis in the future.

NOTE

There are numerous documents detailing 5G projects and use cases. This document describes only the requirements of 5G projects and project scenarios that may become main trends in the future.

2.3.1 Video Transmission

5G network features high-bandwidth and low-latency transmission, promoting the development and application of video services. B2C services focus on cloud VR, cloud gaming, and HD video playback, while ToB services center on HD video surveillance and VR/AR video transmission.

In recent years, HD (4K/8K) live broadcast and video surveillance are becoming increasingly popular and have turned critical enterprise applications with 5G's development.

However, such applications also raise new requirements on the network. If audiences are not close to the video sources, the TCP throughput will not be optimal and the UE-perceived rate will not satisfy the requirements of 4K video stream due to slow start and congestion control mechanism. This will result in decreased quality of experience (QoE). Furthermore, because the video stream upload and download are wirelessly connected, requiring high uplink bandwidth and shorter latency to ensure real-time performance and avoid frame freezing.

Video surveillance in 5G campuses is a special scenario of live broadcast, where download seldom occurs. In this scenario, Terminals collect video in real time and send to the storage server. Therefore, it rather has higher requirements on video quality than real-time performance.However, high requirements on real-time performance is required in industrial applications where video uploaded is used for fault diagnosis to trigger subsequent operations.

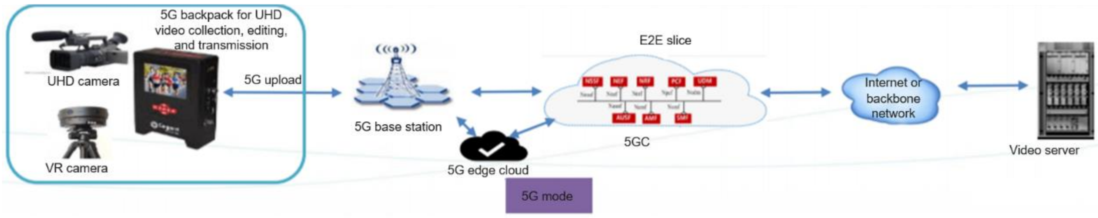
The high efficiency video encoding and decoding (HEVC, also known as H.265) standard reduces the bandwidth requirement by 50% without incurring obvious quality loss. The latest speed improvement in H.265 proves its potential to replace H.264 on 5G networks. H.265 will help alleviate the pressure from the significantly increased bandwidth required by UHD videos. The potential spatial resolution is up to 8K, and the frame rate reaches up to 300 frame per second (fps).

Overall summation:

- Video upload services have different network requirements, especially different latency requirements, depending on application scenarios. The packet loss requirements also vary. For live broadcast services, packet loss will cause artifacts and affect user experience. However, in the scenario where packets are sent back to the server, packet loss triggers the retransmission mechanism, which has little impact on services.
- UDP-based transmission is used to ensure real-time transmission.
- The transmission rate is not the maximum network bandwidth but approximates to the video encoding rate.
- High network stability is required. When the network fluctuates, the instantaneous rate and latency cannot meet the SLA requirements, adversely affecting user experience.
- High uplink bandwidth is required. Congestion occurs if the number of concurrent access services of a base station is greater than the base station capacity. This issue needs to be solved from the perspective of network construction in fixed scenarios. If terminals are mobile, it is difficult to guarantee the network performance in advance. Once terminals move to high-traffic areas, uplink congestions can occur easily.

2.3.1.1 Major Events

Figure 2-3 HD live broadcast networking



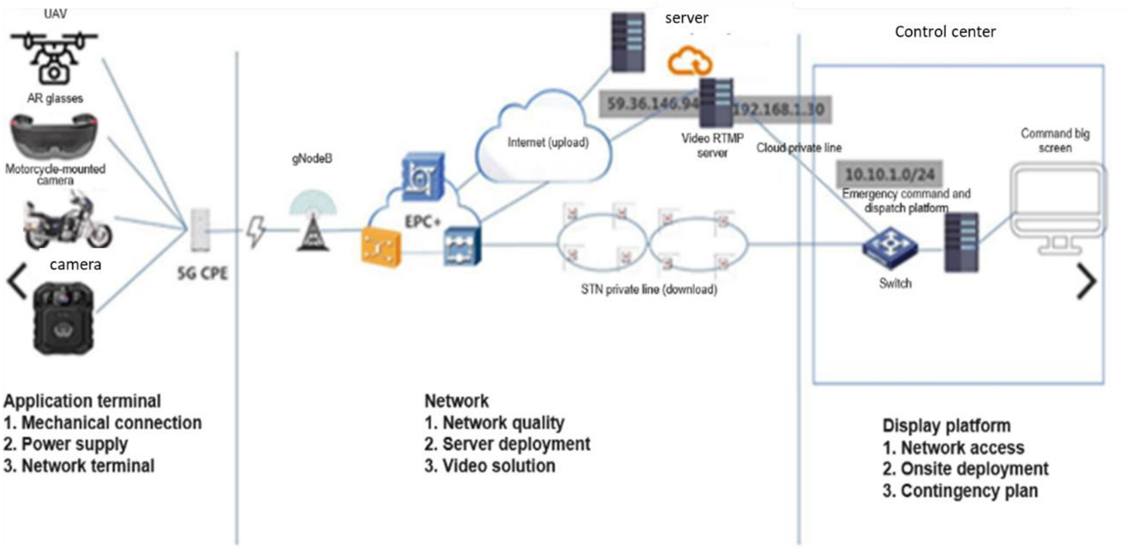
- Live broadcast of project X in China: The live videos are collected through cables and microwave. It is immobile and inflexible. Additionally, the uplink bandwidth of microwave and satellite is insufficient in supporting 4K live broadcast.
- The 5G backpack for ultra-HD video collection, editing, and transmission can be flexibly deployed in areas with 5G coverage. A single cell supports four channels of 4K uplink transmission, meeting requirements in most scenarios. The collection and broadcasting efficiency improves, and the cost is greatly reduced. The solution features flexibility, guaranteed performance, and lower costs.

2.3.1.2 Smart video storage and analysis system

In China, with the development of safe city and smart transportation and enhanced security awareness of users in education, finance, and property industries, the video surveillance market has been growing steadily. However, the majority of cameras still depend on manual monitoring, resulting in an ill-timed processing of video data, poor real-time surveillance performance, and difficult video retrieval. Once an event occurs, video retrieval from massive cameras consumes a significant amount of labor and efforts from people. To solve these problems, the video surveillance industry has been developing and evolving towards HD, network-based, and intelligent video surveillance. The upgrade and innovation of the video surveillance system continuously generate new market demands. Overall, HD videos require a bit rate higher than 1 Mbps, while UHD videos require even higher bit rate, more network traffic, and more storage space. The existing 4G network cannot meet such requirements, and only 5G networks can satisfy ultra-HD videos with a significant amount of data and high real-time performance.

[5G smart video storage and analysis system]

Figure 2-4 Smart video storage and analysis system



Introduction: 5G network capabilities are leveraged to build a 3D smart video storage and analysis system that uses unmanned aerial vehicles (UAVs), AR glasses, motorcycles.

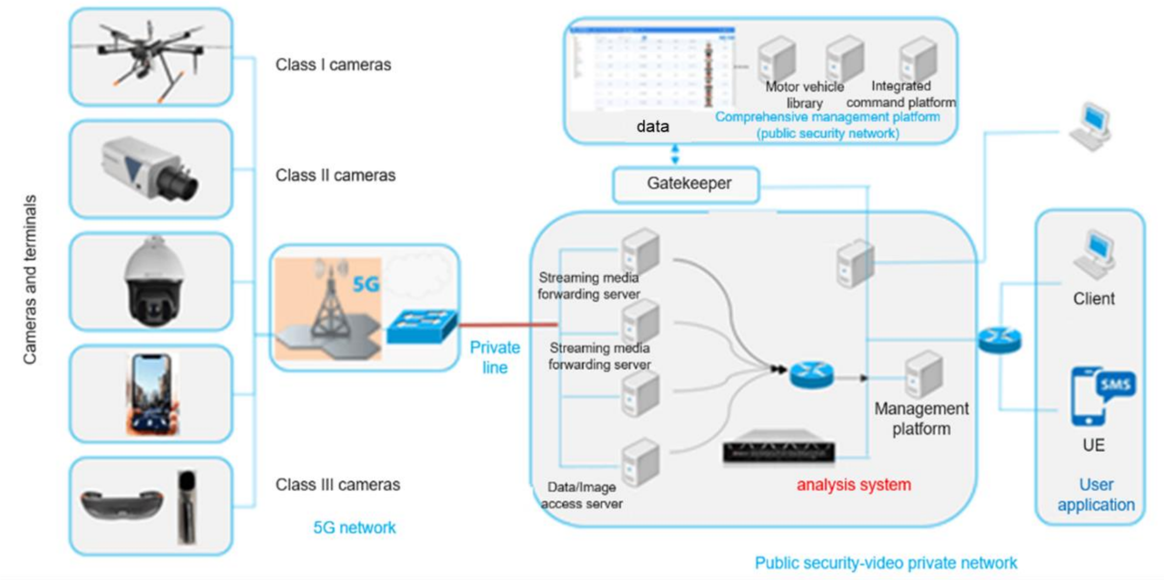
Major application scenarios:

1. Video upload from UAVs: UAVs patrol along preset routes and display object key information.
2. 1080p videos are smoothly transmitted at a frame rate of 30 fps and a bit rate of 8 Mbps.
3. Video upload from motorcycle-mounted cameras: motorcycles equipped with cameras that upload 1080p videos in real time at a frame rate of 30 fps and a bit rate of 8 Mbps. They additionally wear AR glasses to analysis object information. In this scenario, AR does not require heavy traffic.

Network requirements: Terminals currently available on the market do not support 4K video functions, and therefore do not have high uplink bandwidth requirements and 80 Mbps suffices.

[5G smart video record and analysis system]

Figure 2-5 Video surveillance networking



Service introduction: Three types of cameras are connected to the video private network. The analysis server can give object information.

- Wireless network design
  - The gNodeB reads the QoS attribute value of each QoS flow over the N2 interface, when a UE initiates a service request. High-priority services are mapped to high-priority logical channels and area preferentially scheduled.
  - Control plane services are always preferentially scheduled.
- Transport network design
  - Transmission QoS control is performed between the core network and neighboring base stations. Differentiated services code points (DSCPs) are tagged according to data transmission priorities and mapped to VLAN priorities.
  - Determine the priority based on the customer's service requirements or refer to Huawei recommendations.

QoS requirements:

- Ground scenarios

Scenario	Sub-Scenario	Device Quantity	Video Channel Quantity Per Device	Total Video Channel Quantity	Service Type	Uplink Rate	Downlink Rate	Moving Speed
Ground patrol	Video surveillance (fixed pole)	10	1	10	4K	25 Mbps	N/A	0 km/h

Scenario	Sub-Scenario	Device Quantity	Video Channel Quantity Per Device	Total Video Channel Quantity	Service Type	Uplink Rate	Downlink Rate	Moving Speed
	Robot	2	1	2	4K	25 Mbps	N/A	5 km/h
Object recognition	AR glasses	2	1	2	1080p	6 Mbps	N/A	5 km/h
	Mobile phone	2	1	2	2K	10 Mbps	10 Mbps	5 km/h
	car	1	2	2	4K	25 Mbps	N/A	50 km/h
Behavior analysis	Checkpoint/fixed pole	10	1	10	4K	25 Mbps	N/A	0 km/h

- Air scenarios

Scenario	Service Type	Uplink Rate	Downlink Rate	End-to-End Latency (Service)	End-to-End Latency (Control)	Maximum Height	Maximum Speed
UAV	1080p/2K	6–10 Mbps	1 Mbps	<500 ms	<100 ms	150 m	60 km/h
	4K	25 Mbps	1 Mbps	< 200 ms	< 20 ms	150 m	60 km/h

2.3.1.3 Telemedicine

Major application scenarios:

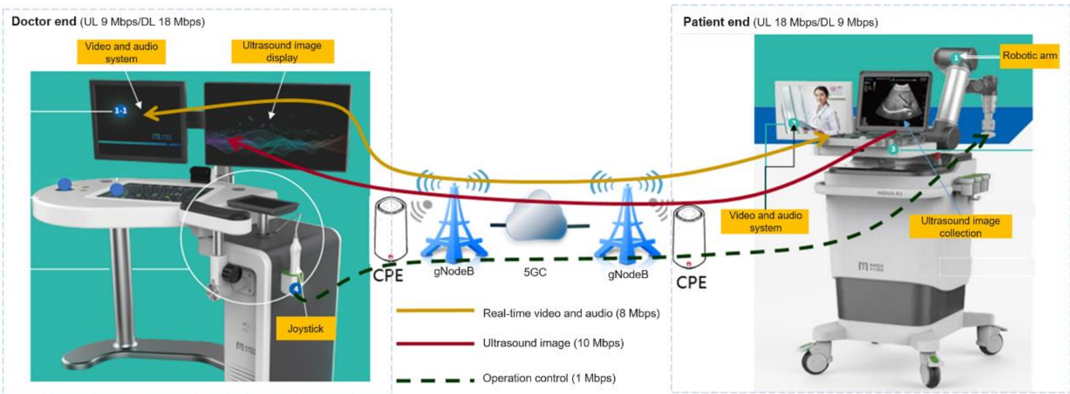
- Remote B-mode ultrasound inspection: The B-mode ultrasound equipment has one end deployed in a hospital for doctor access and one end deployed in a primary healthcare institution for patient access. Doctors can perform diagnosis based on the ultrasound images collected by the remote ultrasound system, and communicate with patients through real-time audio and video.
- Remote emergency: Real-time data, including ambulance location, electrocardiograms, ultrasound images, blood pressure, heart rate, oxygen saturation, and body temperature, is synchronized to the 5G remote command center. Doctors can then diagnose and guide first aiders on emergency treatment through real-time audio and video.
- Remote surgery: Doctors can control robotic arms through their remote desktops and perform remote consultation through cloud video/conferencing.

Network requirements:

Application	Scenario	Patient-End Bandwidth	Latency (ms)
Remote B-mode ultrasound inspection	HD video upload	10 Mbps uplink	100
	Doctor and patient video call	8 Mbps uplink and downlink	100
	Operation control	1 Mbps uplink and downlink	20
Remote first aid	Ambulance video upload	12 Mbps uplink	50
	Video call between ambulance and emergency center	8 Mbps uplink and downlink	100
Remote surgery	Operating table video upload	12 Mbps uplink	100
	Consultation video call	8 Mbps uplink and downlink	100
	Operation control	1 Mbps uplink and downlink	2

Remote B-mode utrasound inspection:

Figure 2-6 Remote B-mode ultrasound inspection networking



Network requirements:

	Uplink	Downlink
Doctor end	9 Mbps (1 Mbps for operation information and 8 Mbps for doctor and patient video communication)	18 Mbps (10 Mbps for ultrasound images and 8 Mbps for doctor and patient video communication)

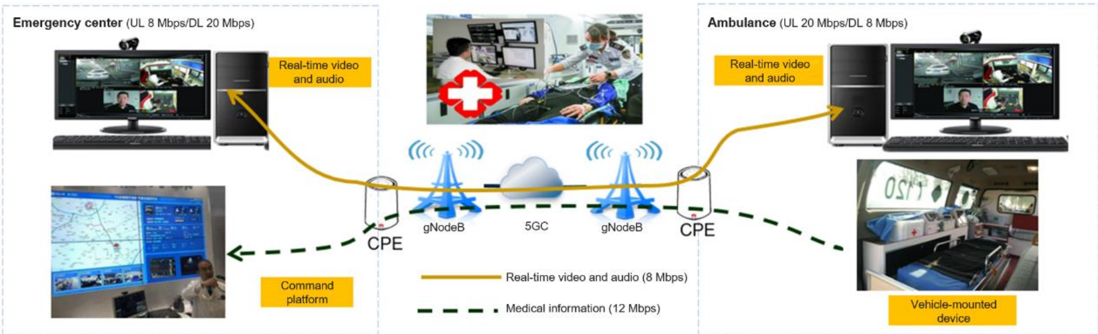
	Uplink	Downlink
Patient end	18 Mbps (10 Mbps for ultrasound images and 8 Mbps for doctor and patient video communication)	9 Mbps (1 Mbps for operation information and 8 Mbps for doctor and patient video communication)

NOTE

The bandwidth is based on 1080p videos and if the video resolution is 4K, the required bandwidth is 25 Mbps.

Remote first aid:

Figure 2-7 Remote first aid networking



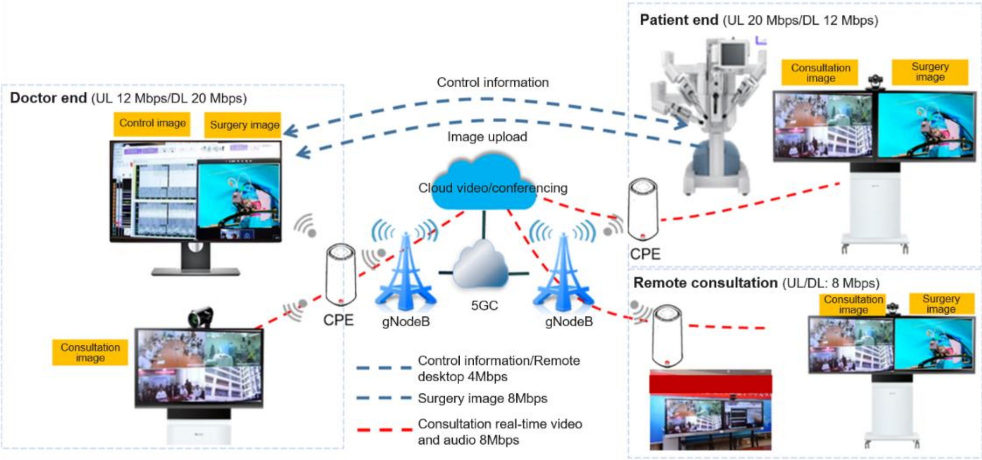
	Uplink	Downlink
Emergency center	8 Mbps (Real-time audio and video)	20 Mbps (12 Mbps for healthcare data and 8 Mbps for real-time audio and video)
Ambulance	20 Mbps (12 Mbps for healthcare data and 8 Mbps for real-time audio and video)	8 Mbps (Real-time audio and video)

NOTE

The bandwidth is based on 1080p videos and if the video resolution is 4K, the required bandwidth is 25 Mbps.

Remote surgery:

Figure 2-8 Remote surgery networking



	Uplink	**Downlink**
Doctor	12 Mbps (4 Mbps for remote desktop and 8 Mbps for consultation video)	20 Mbps (4 Mbps for remote desktop, 8 Mbps for surgery images, and 8 Mbps for consultation video)
Patient	20 Mbps (4 Mbps for remote desktop, 8 Mbps for surgery images and 8 Mbps for consultation video)	12 Mbps (4 Mbps for remote desktop and 8 Mbps for consultation video)
Teleconsultation	8 Mbps (consultation video)	8 Mbps (consultation video)

NOTE

The bandwidth is based on 1080p videos.

Case source: [Use Case] 5G Telemedicine

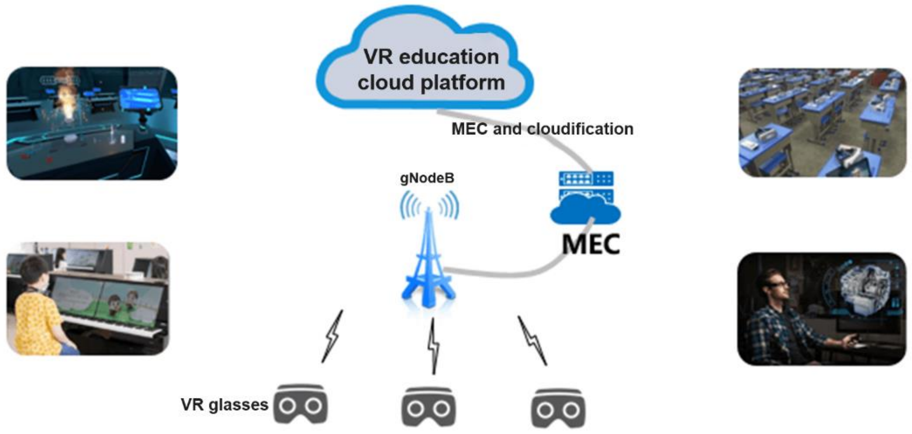
http://3ms.huawei.com/documents/docinfo/1908474

2.3.1.4 Remote Education

There is no significant difference between building a smart school campus and building other campuses. This section describes only the unique service scenarios of remote education.

- AI-assisted teaching
- VR remote teaching

Figure 2-9 Remote education networking



AR/VR teaching contents are uploaded to the cloud. The cloud computing capability is used to implement AR/VR running, rendering, display, and control. The high bandwidth and low latency of 5G are used to transmit the content to VR glasses in real time and construct AR/VR cloud platforms and applications. These include virtual labs, popular science teaching, and 3D interactive classrooms.

AR teaching content example:

VR teaching content example:

Network requirements:

VR teaching:

- Bandwidth: 25–40 Mbps (4K)
- Latency: 80 ms

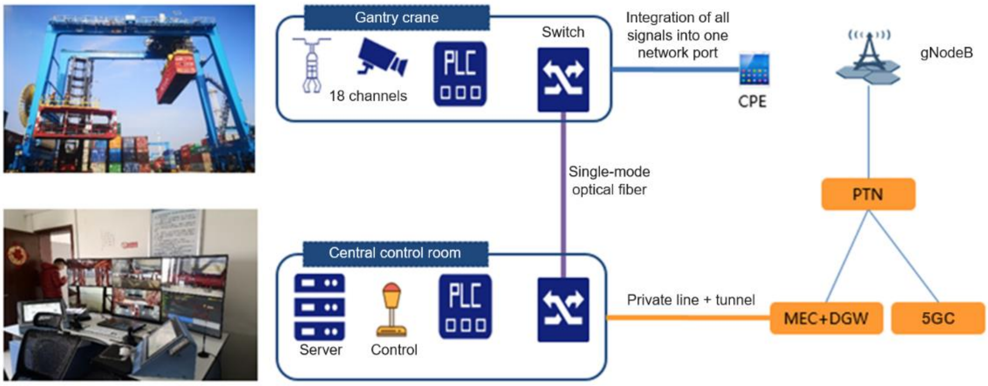
AR Teaching:

- Bandwidth: 25–40 Mbps (4K, not required for most scenarios)
- Latency: 10 ms. Interactive AR operations have high latency requirements.

2.3.2 Industrial Park

2.3.2.1 Smart Port

Smart ports are built through 5G.



Application scenarios: Remote gantry crane control, automated guided vehicle (AGV) control, and autonomous driving of container trucks

- Wireless cameras are deployed on tower cranes and bridge cranes to upload images in real time, enabling personnel to perform loading and unloading remotely in the operation room.
- Remote auxiliary control of AGVs
- 5G-based autonomous driving control of unmanned trucks

Overall, industrial parks are enterprise applications, and their service scenarios are predominantly video surveillance and remote control.

Network requirements:

HD video streams:

- Latency: 50–80 ms
- Bandwidth: 30–100 Mbps
- Reliability: 99.9%

Note: If each crane is installed with 18 channels of HD videos in 1080p and 20 fps, the average media stream bit rate is 2 Mbps, and the required uplink bandwidth is 36 Mbps (18 x 2 Mbps).

Assume that each container yard occupies an area of 450 m x 350 m, each container yard has 14 columns, and each column has two or three gantry cranes. The total uplink bandwidth of such a container yard is 1,510 Mbps (14 x 3 x 36 Mbps).

Crane control signal flow:

- Latency: 18 ms
- Bandwidth: 50–100 kbps
- Reliability: 99.999%

Note: The programmable logic controller (PLC) watchdog signal latency is 18 ms. Ensuring the availability of 18 ms in 5G scearios is extremely important and considerations should be made how to implement it.

2.3.2.2 Commercial Campus

Major service scenarios:

- Industrial HD video surveillance
- Industrial camera: Machine vision to implement intelligent detection of the assembly process
- AR assistance: AR application in the assembly process to provide real-time guidance for employees, and warn and record non-standard operations

Scenario description:

1. 4K video real-time monitoring: Three 4K HD cameras are installed on the patrol car and videos are uploaded to the remote monitoring platform through the customer-premises equipment (CPE).

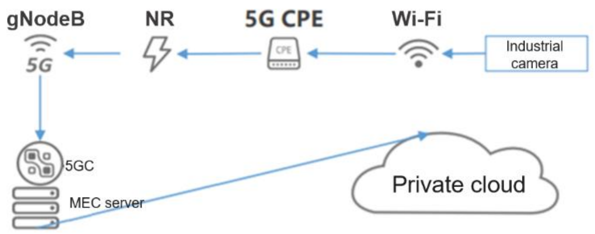
Network requirements:

a) Upload: 25-40 Mbps uplink bandwidth. 75–120 Mbps for three concurrent channels.

b) Download: The maximum peak rate is 1 Gbps, depending on the number of concurrent videos.

2. Industrial cameras: 360-degree photographing and scanning. The 5G CPE network directly sends the scanned images to the private cloud through the mobile edge computing (MEC) for exception detection.

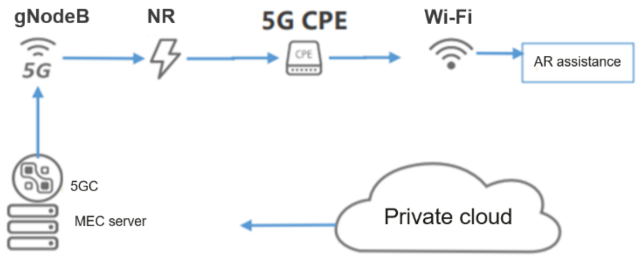
Figure 2-10 Industrial camera networking



Network requirements: Industrial cameras photograph three images of a module per second. Each image is 300–600 KB and seven to eight cameras upload images simultaneously, so 50–115 Mbps uplink rate is required. After photos are combined and processed on the cloud, the final drawing is less than or equal to 500 MB. Onsite engineers need to download the drawing using a tablet or PC within 3 seconds, so 1.3 Gbps downlink rate is required.

3. AR assistance: Image acquisition of the assembly area with the use of AI devices. Assembly site images are uploaded to the cloud in real time through 5G network. The assembly personnel can download related materials from the private cloud on the AR device. They can view the visualized process file information on the AR display in real time to provide real-time guidance.

Figure 2-11 AR assistance networking



Network requirements: Typical AR glasses are mainly used for 720p video streaming, so 1.5 Mbps downlink rate is required and the latency must be within 20 ms. The 5G + MEC solution is required, as 2K–4K video streams are future demands.

2.3.2.3 Smart Fish Pond

Fisheries are predominantly located in remote areas, and major services are monitoring and feeding. Fish-farming is one of the use cases in aquaculture, and the coastal ecosystem is required to be studied as a whole to take advantage of potential 5G network functions.

Customer challenges: Diseases, fish health, predators, food waste, pollution, biomass, fish escaping from farms, extreme/harsh weather conditions, and high costs.

Major service scenarios:

- HD video surveillance
- Remote feeding

Network requirements:

- Uplink: 7.5 Mbps for each camera and 75–135 Mbps for each fishery
- Video surveillance latency: The maximum latency is 0.5s, offering complete video experience for end users.
- Ping latency: Around 30 ms now and around 10 ms by 2021
- Interaction latency: 200 ms (round-trip)

2.3.3 Industrial Automation

2.3.3.1 Smart Grid

Based on the high-speed bidirectional communication network, advanced sensing and measurement technologies, control methods, and decision-making support systems are used to achieve intelligent power grids throughout the power generation, transmission, transformation, distribution, and consumption phases.

Low-voltage centralized metering is a typical mMTC service.

Table 2-1 Smart grid network requirements

Service	**Communications Requirements**		**Typical Power Terminal**	**5G Communications Terminal**	**Communications Terminal Quantity**	**5G Network Slice**	**Special Network Requirements**
	Latency	**Bandwidth**					
Differential protection for smart distribution networks	≤ 15 ms	≥ 2 Mbps	Intelligent DTU	CPE	6	URLLC	Clock synchronization: < 10 μs
Automatic "three-tele" services for power distribution	≤ 50 ms	≥ 2 Mbps	Intelligent DTU	CPE		URLLC	N/A
Automatic "three-tele" services for power distribution	≤ 50 ms	≥ 2 Mbps	Intelligent DTU	CPE	6	URLLC	N/A
Power grid emergency communications	≤ 200 ms	20–50 Mbps	UAV, mobile phone, and camera	Hub and CPE	4	URLLC eMBB	N/A
Low-voltage centralized metering	≤ 3 s	1–2 Mbps	Concentrator and meter	Customized communications compartment	100	mMTC	Massive connection
Precise load control	≤ 50 ms	1.13 Mbps	Intelligent DTU	CPE	6	URLLC	High reliability Low latency

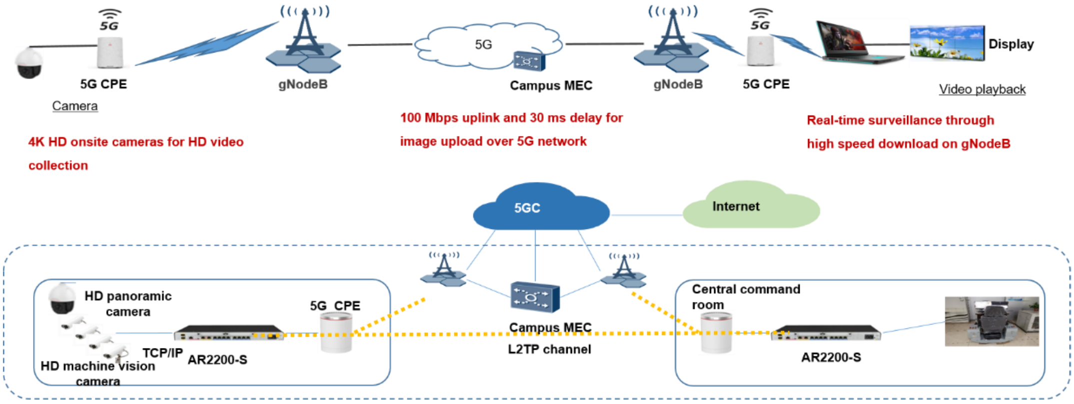
Major service scenarios: eMBB, URLLC, and mMTC. However, smart grid encounters great challenges in mMTC scenarios.

2.3.3.2 Smart Iron and Steel Plant

Crown block precise control requires remote zero-wait and short latency, as well as high definition, high precision, and multi-view UHD video signal switching.

Service Scenario 1: Real-time HD video surveillance in high-risk areas and harsh environments

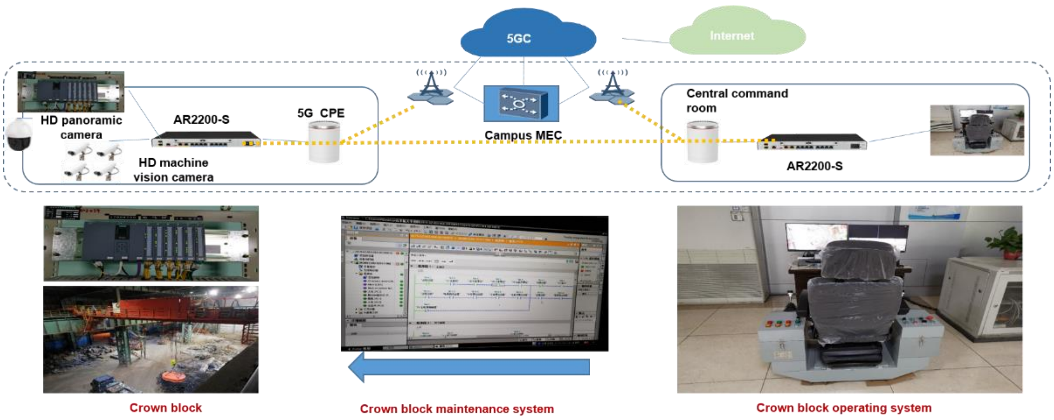
Figure 2-12 HD video surveillance networking



The control center can accurately learn about the onsite situation through real-time HD video surveillance, generate warnings in real time, and intervene in advance to avoid safety misadventure. This additionally reduces patrol workload, improves efficiency, and prevents production accidents.

Service scenario 2: PLC-based remote video control in high-risk areas and harsh environments

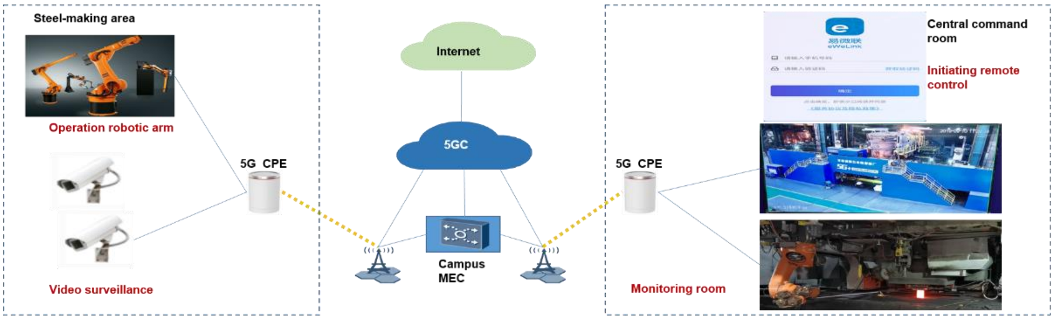
Figure 2-13 PLC-based remote control networking



The solution consists of the crown block operating system, 5G network, and crown block (including PLC and camera). The crown block operating system is used to remotely control the crown block in real time. The 5G's low latency feature offers operators with HD videos from the first angle of view and enables zero-latency control, ensuring precise and real-time remote control. It frees operators from noisy, dusty, and high-temperature environments. The feature further improves the working environment, raises work efficiency, and prevents production accidents.

Service scenario 3: Remote robotic arms in high-risk areas and harsh environments (on-click slag addition)

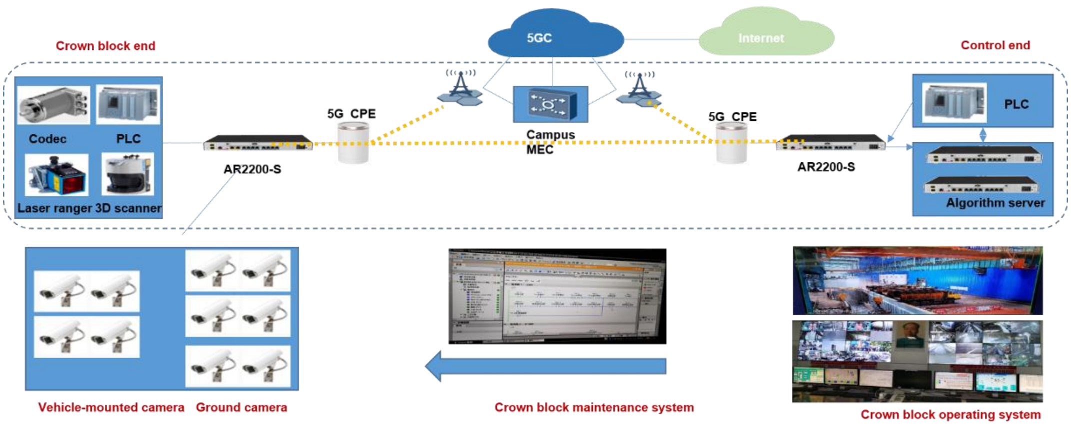
Figure 2-14 One-click slag addition networking



The robotic arm can operate independently through one-click remote control. The slag-adding mechanical arm can be remotely connected to the control system through a 5G mobile phone to implement remote one-click control anytime and anywhere. The mechanical arm next to the high temperature boiler can run automatically, and the slag-adding mechanical arm sprays the iron ore slags evenly into the steel-making boiler to improve the steel production quality. This prevents workers from working near high temperature boilers, improves the working environment, reduces labor costs, and avoids production safety hazards.

Service scenario 4: Unmanned crown block

Figure 2-15 Unmanned crown block networking



The unmanned crown block system consists of collectors (including 3D scanner, laser ranger, coderc, and camera), 5G network, and PLC. The scanner collects information about the horizontal and vertical directions. The laser ranger collects distance information, and the camera collects information and images about surrounding materials, pits, vehicles, and bucket height as well as loading and unloading positions, and transmits data to the MEC (to be deployed later) in real time for data processing and establishing onsite data 3D models. At the same time, AI algorithms construct action instruction sets and deliver them to the crown block for execution, implementing unmanned crown block for production.

5G network overall requirements:

Table 2-2 Video surveillance network requirements

Application Scenario	Device Quantity	Bandwidth Requirement
Video upload	M cameras (1080p) and one panoramic camera (4K) on each crown block	1080p: 4 Mbps per channel, N x 4 Mbps per crown block 4K: 32 Mbps per channel, 32 Mbit/s per crown block

Table 2-3 Network requirements on PLC-based remote video control

Applicaion Scenario	Device Quantity	Bandwidth Requirement	RTT
Remote control	One PLC module on each crown block	/	20-50 ms

3 For details about the experience requirements and baselines, see section 4 5G ToB Service Modeling Framework

3.1 ToB & B2C Modeling Differences

Table 3-1 B2C and ToB modeling analysis dimensions

Analysis Dimension/Service Type	B2C Experience Modeling	ToB Experience Modeling
Quality commitment	Operators do not make SLA commitments to individual users.	Operators make SLA commitments to enterprise customers.
Satisfaction	Complaints and churn occur following poor single-user experience: Individual experience modeling is important.	No complaints from things but if the overall SLA does not meet the requirements, customers may claim for compensation based on contracts. Group experience modeling is more important.
Troubleshooting	B2C users can rectify the fault by themselves by powering off, calling the assistance hotline, or consulting associates.	ToB users cannot. Once a fault occurs, the system will be suspended.
Traffic model	B2C services are bursts (long-time or short-time data transmission).	ToB services are continuous or involve regular bursts. Services are always online.

Analysis Dimension/Service Type	B2C Experience Modeling	ToB Experience Modeling
Software application	B2C software systems are dominated by few providers.	The ToB-oriented software systems present long tail characteristics.
Network characteristics	Shared networks cannot guarantee differentiated QoS for different services.	SA networking and slicing: ensures differentiated QoS requirements of different services.
Technical challenges	Encrypted identification, experience modeling, and traffic explosion	E2E QoS measurement (UDP) and dynamic QoS guarantee
KPI difference	Throughput and latency	In addition to throughput and latency, consider energy consumption, network resource usage, and abnormal distribution of objects.
Optimization points	Wireless RF quality, CN-SP route/rate limiting/packet loss	Wireless RF quality, network structure adjustment, and resource allocation policy

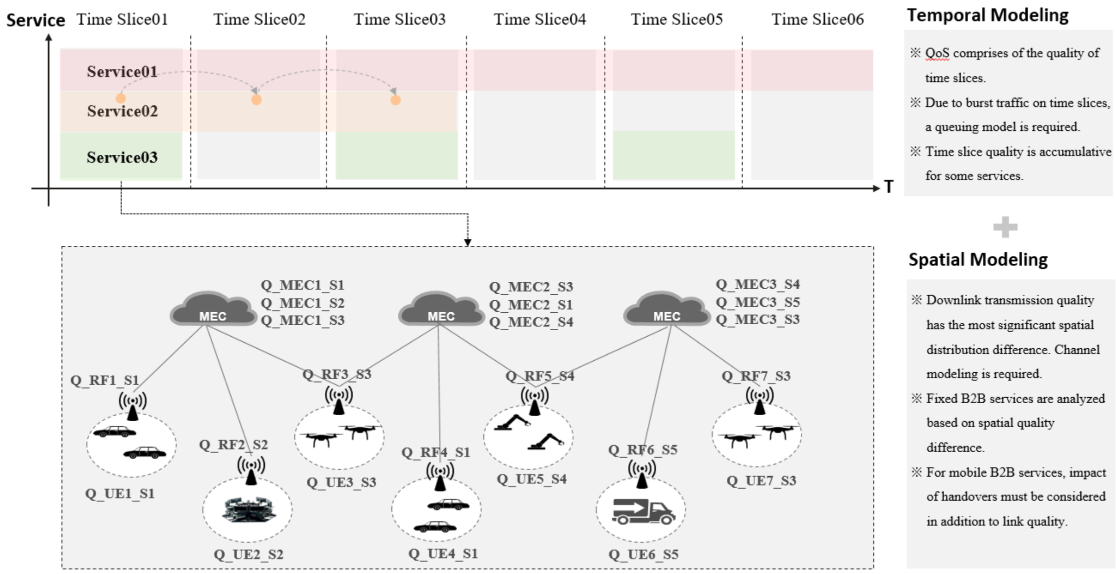
Table 3-2 B2C and ToB modeling differences

B2C Experience Modeling	ToB Experience Modeling
PSPU individual experience modeling	Group quality modeling for "things"
Ensuring the monopolistic applications, categorized experience modeling	Too many application scenarios, customized modeling + general categorized modeling
Strive for ultimate user experience	Optimal balance between network resources and experience
Precise QoS measurement (for example, RTT 99.9% precision)	Precise measurement + AI-based quality prediction (with confidence)
Experience evaluation and demarcation are the driving force of service solutions.	Experience assurance is the driving force of service solutions.

3.2 ToB Modeling Method Exploration

3.2.1 Fine-grained Spatio-temporal Modeling

Figure 3-1 ToB QoS and space-time relationship



ToB QoS modeling must be based on time and space.

- In the time domain, event-driven instantaneous traffic impact or packet quantity impact causes burst latency and packet loss. As radio devices and core network/bearer network devices perform instantaneous queuing, the queuing mechanism of different services needs to be modeled to quantify the QoS impact such as the latency, packet loss, and bandwidth of service processing within a time slice.
- In the space domain, rapid changes in channel quality caused by mobility and communication location changes are quantified in channel quality modeling. MEC resource allocation also impacts the overall ToB service quality. If MEC planning and resource allocation are properly performed, the MEC processing latency and application layer processing latency are not closely related to the location. However, this parameter is not considered.
- Spatio-temporal modeling: First, the quality of a single time slice is modeled, with the impact of the queuing model and the channel model on the quality fully considered in the single time slice. For services that span multiple time slices, check whether the quality of the time slices is related. If they are related, use the state change method (such as Markov) for associated evaluation. If the quality of time slices is independent, the weighted average or the moving weighted average considering the near-end effect can be used to evaluate the comprehensive service quality.

The average performance alone is insufficient for ToB service quality evaluation. Transient poor quality (burst latency, burst congestion, burst packet loss, and burst jitter) is generated at the time slice level. Different objects in the same time slice represent different positions in space and may change over time. Assume that a location of an object in a same time slice is fixed. It should be noted that service quality varies greatly at different locations, that is, the final comprehensive quality is a two-dimensional function of time and space.

The following uses end-to-end latency as an example. Assume that the latency (D) is a two-dimensional function of time and space:

$$D(x_i, y_j), i \in [1, N], j \in [1, M]$$

In the formula, x_i is the current time slice, the maximum quantity of time slices in the evaluation period is N, y_j is the current object, and each object has a different location, thus represents a different spatial location, and the maximum quantity of objects is M.

$$f(x, y) = \iint_{i=1,j=1}^{i=N,j=M} D(x_i, y_j)dxdy$$

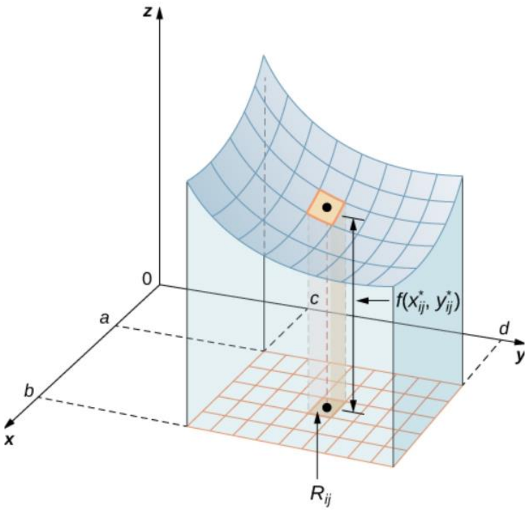
$f(x, y)$ is the double integral of D in the space-time dimension and represents the accumulated measurement of the latency, which is presented as the volume of the blue area in the figure.

$$f^*(x, y) = \iint D^*(x_i, y_j)dxdy\,, D^*(x_i, y_j) > D_{boundary}$$

$f^*(x, y)$ indicates the sample integral of the latency that exceeds the boundary.

$$r(x, y) = P(D > D_{boundary}) = \frac{f^*(x, y)}{f(x, y)} \times 100\%$$

Figure 3-2 Double integral representation of spatial-temporal distribution of mass



In a measurement period, the E2E latency of all objects can be measured using $f(x, y)$ and $r(x, y)$.

3.2.2 Scenario-based Event-driven Modeling

In ToB networks, wireless sensors are used frequently, and all applications that use in-situ sensors strongly depend on their proper operation, which is difficult to ensure. These sensors are usually cheap and prone to failure. For many tasks, sensors are used in harsh weather conditions, making them more vulnerable to damage. In addition, industrial devices have high requirements on reliability. Common faults can be detected by alarms in the tenant system. However, hidden faults are difficult to detect due to external factors or aging. If they are not handled in a timely manner, faults gradually occur, reducing the SLA. Therefore, the pre-detection and pre-analysis of the abnormal behavior of objects are significant to the preventive management of enterprises. And because of the number and variety of things, the detection process must be automated, scalable, and fast enough for real-time streaming data.

In conclusion, machine learning and heuristic learning-based anomaly detection technologies will play an increasingly important role in various future 5G IoT applications.

Anomaly Detection in Intelligent Inhabitant Environment

In intelligent inhabitant environment, embedded sensor technology plays a major role in monitoring occupants' behavior. The inhabitants interact with household objects, and embedded sensors generate time-series data to recognize performed activities. Generated sensor data is very sparse, because the sensor values change when the inhabitant interacts with objects. The need for robust anomaly detection models is essential in any intelligent environment.

- Statistical methods in intelligent inhabitant environment

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Dissimilarity Measures [17]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	State and motion sensors	Smart home	Distance index
Percentiles [18]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Motion and door sensors	Senior's inhabitant home	False alert rate and sensitivity
Gaussian Mixture Model [19]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Occupancy, motion and door sensors	UK sheltered housing scheme	Histogram Visualization
Hierarchical Markov Model [20]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Embedded state sensors in Home	MIT Placelab smart home	Accuracy
Switching Hidden Semi-Markov Model [21]	Continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Video cameras	Smart kitchen	Average Accuracy
Bayesian Model [22]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Binary sensors, motion sensors and pressure sensors	Smart home	Precision, Recall, and F-measures

- Machine learning methods in intelligent inhabitant environment

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Single Class Support Vector Machine (SVM) [16]	Binary data	Collective anomaly	Supervised learning over the patterns of normal behavior	State sensors deployed in living room, kitchen and dining area	Smart home	Type I and II error, Precision, Recall, F - Measure
Multi-class SVM [24]	Continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Accelerometer and gyroscope sensors	Wearable gadgets	Accuracy, Precision, Sensitivity
Support Vectors [25]	Continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Infrared motion sensors	Ubiquitous Healthcare House	Positive predictive value
Principal Component Analysis (PCA) and Fuzzy Rule-based System [26]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	state and motion sensors	Smart home	Hotelling's T2 and Square prediction error (SPE)
Kernel Nonlinear Regression and SVM [27]	Continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Light, temperature, microphone, accelerometer, magnetometer	Human activities	False positive rate
Convolutional neural network (CNN) and Recurrent Neural Network (RNN) [28]	Continuous data and Images	Collective anomaly	Supervised learning over the patterns of normal behavior	Microwave sensor and video camera	Smart home	Accuracy and Mean Absolute Error (MAE)

Anomalous Behavior in Intelligent Transportation System

- Statistical methods in intelligent transportation systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Kernel Feature Space and PCA [30]	Multi-dimensional time-series	Collective anomaly	Supervised learning over Telemetry data from space station	System components (Soft sensors) in spacecraft	Aerospace	False alarms
K-means and GMM [31]	Numerical, categorical, and binary.	Contextual anomaly	Unsupervised learning over EuroFOT database	403 parameters of hardware and soft-ware sensors,	Anomaly detection for road traffic	Colorscales (red for anomalies)
Structured sparse subspace learning [32]	Continuous data.	Contextual anomaly	Supervised learning on Thor Flight 107 and 111 flight data	flight-critical sensors	Anomaly detection for flight safety	Accuracy and ROC

- Machine learning methods in intelligent transportation systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Decision tree (C4.5) and Fu-sion Model [33]	Unstructured data in the form of reports	Collective anomaly	Supervised learning over historical data collected over 5 years	Operational data of engines of A-320	Aerospace	Accuracy and Error rate
Multiple Kernel Learning [34]	Discrete and continuous data	Collective anomaly	Supervised learning over the pattern of flights	Different 160 pa-rameters of FOQA dataset	Aviation safety	Accuracy
Extreme Learning Machines [35]	Continuous data	Collective anomaly	Supervised learning on 43000 flights data	Radar, aircraft trajectories and nearby aircraft distance	Aviation safety	Area under the curve
Reinforcement Learning [36]	Continuous data	Point anomaly	Supervised learning on drone captured data	Temperature sensor	Unmanned aerial vehicles	Accuracy
Regression Model [37]	Continuous data	Contextual anomaly	Supervised learning on historical data	Software Logs	Air Traffic Control systems	Precision, Re-call and Ac-curacy
Support Vector Machine [38]	Continuous data	Contextual anomaly	Supervised learning on historical data	GPS and Accelerometer	Air Driving Patterns and Road Anomalies	Accuracy
Deep Convolutional Neural Network [39]	Image data	Point anomaly	Supervised learning on CTIV platform used to collect the images	Camera	Railway track	Accuracy
Autoencoders [40]	Continuous data	Point anomaly	Unsupervised learning on data from B777-200 of civil plane	Gas turbine engine	Aerospace	Precision, Re-call, F1 Score

Anomalous Behavior in Smart Objects

The smart object is a fast-growing area to connect multiple objects together and enable communication between them. It collects valuable data that can be a source of information and knowledge for a wide range of applications. During our research, we found the following statistical and machine learning literature that is aligned with our research questions and search criteria

- Statistical methods in smart objects

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Threshold Level [41]	Continuous data	Point anomaly	Supervised learning on patterns of the food in trash bin	Gas sensors	Trash bin	No performance measures
Analysis of variance (ANOVA) [42]	Continuous data	Collective anomaly	Supervised learning on the driving section	Vehicle data from engine, fuel, gear and steering wheel.	Abnormal behavior of the vehicle	Normal vs abnormal data
Latent Correlation Probabilistic Model [43]	Continuous data	Collective anomaly	Supervised learning on patterns of concrete trucks	Engine speed and pump speed	Concrete pump trucks	Precision, Recall, area under the ROC
Expectation Maximization [44]	Binary data	Point anomaly	Unsupervised learning on switches patterns	Light switches	Smart nursing homes	False alarm rate

- Machine learning methods in smart objects

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Rule-based System [46]	Continuous and image data	Collective anomaly	Supervised learning	Temperature, humidity, light intensity, audio and image data	Efficient energy consumption in school building	Number of conflicts and Miss detection rate
Belief-rule-based Association Rules [45]	Continuous data	Collective anomaly	Supervised learning on patterns of incomplete and vague data	Temperature and rain gauge	Flood prediction	Area under the curve
Cluster heat maps [48]	Continuous data	Contextual anomaly	Supervised learning on patterns of smart cities IoT data	Measurement of total power, reactive power, phase angle, light, temperature, motion sensor, noise level, and vibration	Smart City	Normal vs abnormal comparisons
Temporal Clustering Technique [47]	Continuous data	Collective anomaly	Supervised learning on 8200 parking data	On-street parking spaces and gate controlled sensors	City parking of San Francisco	Correlations observations
Spatio-temporal Framework [49]	Continuous data	Collective anomaly	Supervised learning over 50 devices	Temperature, humidity sensor	Environmental data analysis of Taipei	Accuracy
Features Extraction and Visualization [50]	Continuous data	Collective anomaly	Supervised learning on patterns of smart cup usage	Raspberry Pi Zero, 9-DoF IMU, liquid level sensor, and force sensing resistors	Smart cup	No comparison

Anomalous Behavior in Healthcare Systems

Anomaly detection, analysis, and prediction are considered a revolution in redefining health care systems. In such systems, a clear impact can be seen on health management and wellness to improve quality of life and remote monitoring of chronic patients. Such systems pose a great challenge to reducing the generation of false alarms. In our systematic literature survey, we have found sufficient approaches and methods to identify anomalous behavior of sensors, humans, or machines in healthcare environment.

- Statistical methods in healthcare systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
RMS Graph [51]	Continuous data	Contextual anomaly	Supervised learning on wearable patterns	Accelerometer	Seizure complexities	Threshold configuration
Dynamic Time Warping (DTW) and Density Functions [53]	Continuous data	Contextual data	Supervised learning on Physionet PPG signals	Photoplethysmogram (PPG) signals	Cardiac behavior	Precision, Recall, Specificity
Autoregressive integrated moving average (ARIMA) [56]	Textual data	Point anomaly	Electronic health records	Electronic health records	Clinical decision support system	No performance measure
Hidden Markov Model [54]	Continuous data	Contextual anomaly	Supervised learning on eWALL Consortium parameters	Wearables Fitbit	Sleep analyzer	No performance measure
Hidden Markov Model [52]	Continuous data	Point anomaly	Supervised learning on glucose level	Medical devices insulin tolerance test	Blood glucose level behavior	Precision and Recall
Spectral Coherence Analysis [55]	Continuous data	Point anomaly		Accelerometer Sensor	Gait freezing in Parkinson's disease	Sensitivity

- Machine leaning methods in healthcare systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
HEAL Model [57]	Binary data	Contextual anomaly	Supervised learning on smart home data	Tap sensors on the objects	Abnormal behavioral	No performance measures
Incremental learning algorithm [58]	Continuous data	Contextual anomaly	Supervised learning on the dataset from UCI repository	Heart rate and blood pressure measuring sensor	Anomaly pattern in the health records	Actual vs predicted values
Transductive Transfer Learning [60]	Continuous data	Contextual anomaly	Semi-supervised learning on arrhythmias dataset	ECG monitoring sensor	Electrocardiogram abnormalities	G-mean
A Graph-based Approach [59]	Textual data	Contextual anomaly	Supervised learning on Hospital dataset	Access logs	Anomalies in electronic medical record	ROC curves
Support Vector Machine [61]	Textual anomaly	Contextual anomaly	Supervised learning on EHR	Data logs	Patient-management actions	True alert rate

Anomalous Behavior in Industrial Systems

In industrial systems, the design and development of anomaly detection methods are crucial to reduce the chance of unexpected system failures. It has been found that the developed methods for anomaly detection have been successfully applied to predictive and proactive maintenance. Such methods are widely used to improve productivity performance, save machine downtime, and analyze the root causes of faults.

- Statistical methods in industrial systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Temporal Logic [62]	Continuous data	Point anomaly	Supervised learning on 50 collections of traces	air compressor motor speed	Fuel cell vehicle	Misclassification rate
Correlation Analysis [63]	Continuous data	Contextual anomaly	Supervised learning on 5 machines data	Tap sensors on generator	Electric generators in factories	Correlation coefficients
Density Function Model [64]	Continuous data	Point anomaly	Supervised learning on 24 solar panels data	Electric current data	Solar power generation systems	ROC curves
Markov chain [65]	Continuous data	Point anomaly	Supervised learning on pressure sensor data	Pressure sensor	Oil pipeline	Accuracy

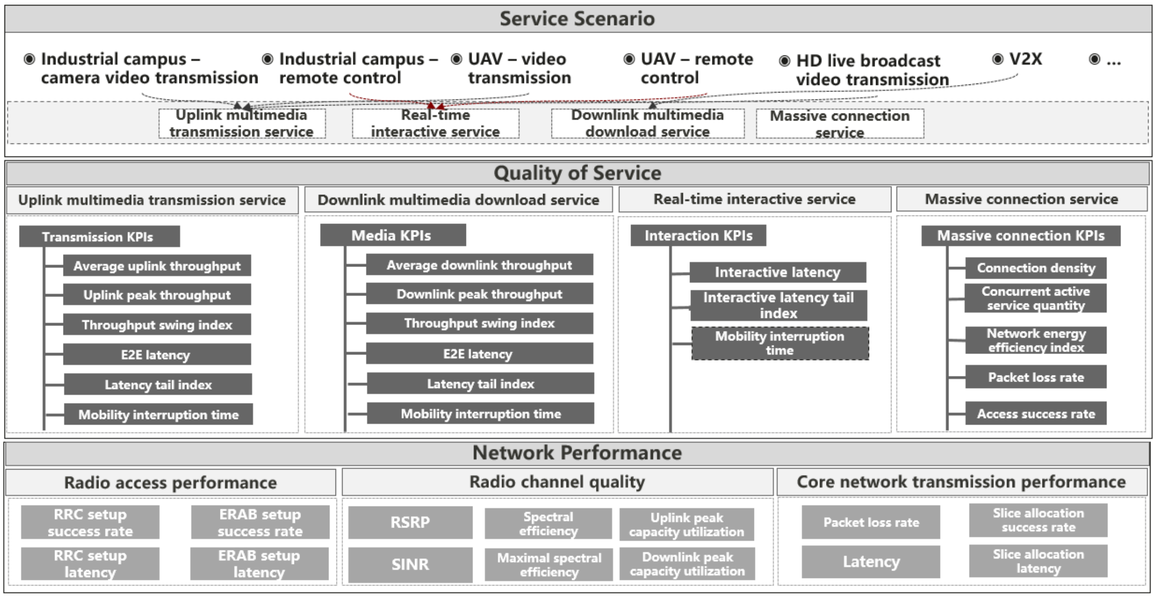
- Machine leaning methods in industrial systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Extreme Learning Machines [69]	Continuous data	Point anomaly	Supervised learning on combustor chambers exhaust data	Temperature sensor	Power plant operations	ROC curves
Multivariate Clustering [71]	Continuous data	Contextual anomaly	Supervised learning on real-world sensory data	Sensor data from the electrical, water, and gas systems	Reduce electricity waste	Misclassification rate
Clustering [67]	Continuous data	Contextual anomaly	Unsupervised learning on five floor building	Temperature sensor	HVAC System in Smart Buildings	False alarms
One-Class SVM [66]	Continuous data	Point anomaly	Camshaft revolutions	A generic DAQ card (NI-6143)	An industrial seal machine	
Neural Network [68]	Continuous data	Contextual anomaly	Supervised learning on the data of thermal power plant	Superheated steam temperature, flow and steam cooling water flow.	Thermal power plant	Room mean square error
Conditional Gradient boosting decision tree (GBDT) [70]	Continuous data	Contextual anomaly	Supervised learning on the data of wind turbine	150 measurement parameters of wind turbine.	Wind turbine	Accuracy

3.3 ToB Modeling Frame

3.3.1 Indicator-driven Modeling Framework

Figure 3-3 ToB indicator-driven modeling framework analysis



ToB service modeling framework:

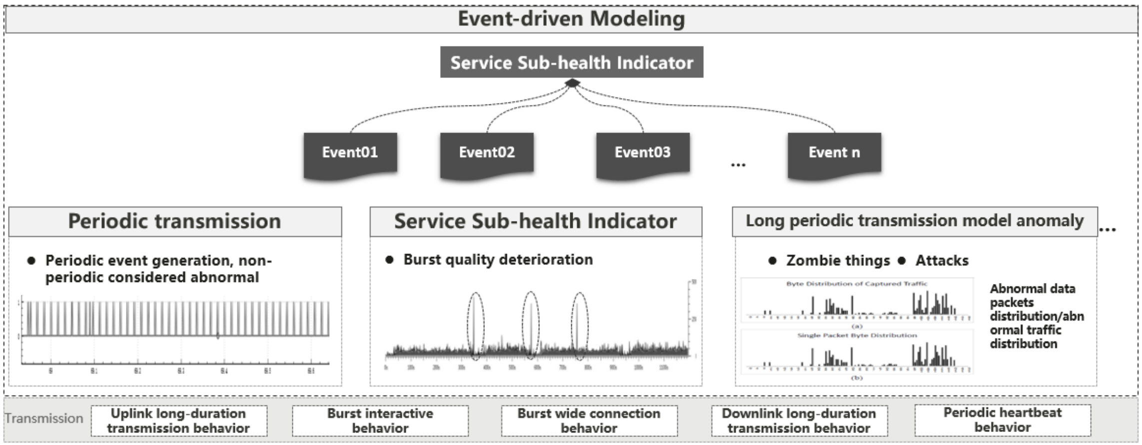
- Service scenarios: The various ToB service scenarios can be generally categorized into three types: eMBB, uRLLC, and mMTC. Based on the understanding of service requirements of the current 5G project, from the perspective of network requirements, the services can be classified into uplink multimedia transmission services, downlink multimedia transmission services, real-time interactive services, and wide connection services. Currently, high-mobility services are not seen in projects. Theoretically, high-mobility services are a special scenario of multimedia services and real-time interactive services.

- Service quality: Different service types have different requirements on networks. Differentiated indicator systems are recommended.
- Network performance: Network is the foundation of service quality. Network performance can be classified into radio access performance, radio channel quality, and core network transmission performance.

Due to radio resource preemption and resource insufficiency, neither 4G nor 5G networks can meet each user's quality requirement. The SLA is not specific to individual users, rather, it is specific to the entire network.

3.3.2 Event-driven Modeling Framework

Figure 3-4 ToB event-driven modeling framework



This document describes the event-driven modeling method, which touches upon problems that cannot be covered by traditional latency/rate indicators. This method is from the perspective of identifying anomalies and analyzing problem types through big data and clustering analysis, rather than the perspective of PSPU experience modeling. To indicate the overall poor quality of a service behavior, you can set the service sub-health index as the comprehensive service quality evaluation indicator affected by anomalies.

General network data transmission behaviors can be classified into the following types:

- Uplink long-duration transmission behavior
- Downlink long-duration transmission behavior
- Burst wide connection behavior
- Burst small packet interaction
- Periodic heartbeat behavior

There is no universial anomaly detection solution that can help define the pattern of anomalies based on the service scenario and possible symptoms. Instead, case-by-case definition of anomaly events based on service characteristics is required. For development of platform products, consider the template, invoking mechanism, and upper-layer statistical indicator calculation model defined for abnormal events, which can be fixed to form product capabilities of edge and central nodes. However, behavior analysis and anomalies definition based on service behavior analysis by service delivery experts are required in frontline projects.

4 5G ToB Service Indicator System

4.1 Uplink Multimedia Transmission Service

4.1.1 Impact Factor

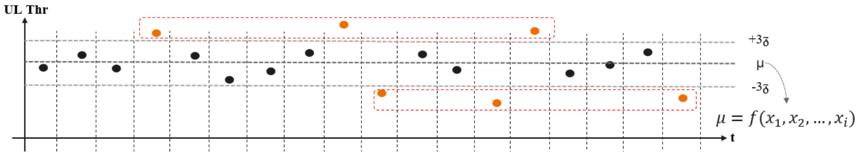
Multimedia transmission services include the following types:

1. For uplink real-time streaming media transmission services, such as live stream download and video surveillance, videos are transmitted in real time based on the video quality, such as the frame rate, bit rate, and resolution. The network transmission rate must meet the bit rate requirements, while the transmission latency must meet certain requirements.
2. Uplink multimedia message services, such as voice messages, picture messages, and video messages are transmitted to the server. Such non-real-time transmission services are generally one-off best-effort transmission. The higher the rate requirement, the better the service experience. The service experience mainly depends on the transmission latency, which is closely related to the rate and file size and is not an objective indicator. The uplink rate is used as the core evaluation indicator.

Table 4-1 Core factors affecting the uplink real-time streaming media transmission services

Factor	Impact
Bit rate	The bit rate refers to the number of audio or video bits transmitted or processed per unit time. It is a common indicator for measuring the audio and video quality. Specifically, a high resolution, high frame rate, and low compression rate usually lead to an increase in bit rates given the same coding used.
Frame rate	The frame rate indicates how frequently pictures appear on display continuously in the unit of frame. The frame rate of the video content must be compatible with the frame rate attribute of the display device. For example, live broadcast services have higher requirements on frame rate stability. Frame rate fluctuation may deteriorate the quality of transmission videos in live broadcast.
Resolution	The video resolution indicates to the number of pixels contained in the video content. The video resolution must be compatible with the resolution of the display device. Otherwise, the video resolution may decrease or the video may not be displayed. For real-time live broadcast transmission services, the resolution is fixed and is closely related to the capabilities of terminals (such as cameras).
Packet loss	Packet loss has a significant impact on the quality of multimedia content. When a packet including an I-frame is lost, all subsequent frames of a same GOP frame depending on the frame are lost, which, as a result, could cause pixelization, frame blocking, and video output stalling. This can also be applied to audio streams. Packet loss can be measured by the average packet loss rate or burst packet loss rate. Burst packet loss has a greater impact on the system. Therefore, it must be considered separately.
Data packet latency	When a packet is transmitted from the source to the destination, transmission latency occurs. If the latency reaches a certain threshold, image blocking and image damage may occur.

Factor	Impact
Jitter	The propagation latency is not constant in a period of time. Therefore, the latency is changing across the entire network. Jitter is an indicator to measure this variability. The real-time streaming media service requires stable IP streams. Jitter may cause buffer overflow and underload, resulting in pixelization and frame freezing of the streaming content.
Average throughput	Rate is a key indicator for ensuring video transmission quality. The average rate alone is not enough. For real-time transmission services, rate fluctuation causes buffer overflow during video transmission. As a result, video frames cannot be played smoothly.
Peak throughput	Peak throughput in the real-time streaming media transmission service is measured by the bit rate when the quality is high. In fact, the peak throughput usually does not reflect the network transmission capability. Therefore, the peak throughput is the maximum throughput that can be reached instantaneously. This number reflects the transmission performance of the pipe.

Throughput swing	Throughput swing is defined as the proportion of the throughput that exceeds that of the previous or next session. It indicates the throughput fluctuation.
	
	The average performance cannot reflect the uplink multimedia transmission service experience. Burst congestion or deterioration will adversely impact user experience.
	● The "dynamic index" and "swing index" are introduced to reflect the fluctuation.
	  ● The proportion of the upward fluctuation that exceeds the range of μ+3σ is the upward swing index.
	  ● The proportion of the downward fluctuation that exceeds the range of μ-3σ is the downward swing index. It is the most essential indicator of quality deterioration.
	  ● For uplink real-time transmission services, a lower swing index indicates stabler transmission performance and better user experience.
	● From the perspective of real-time measurement, the values of μ and 3σ are calculated based on the average value and variance of the current time. Therefore, their values change dynamically. In the figure, μ and 3σ are represented as an f(x) curve.

Mobility interruption time	Mobility interruption time refers to the service interruption latency generated when a terminal moves. This indicator does not apply to fixed terminals.
Interactive latency	Interactive latency indicates the latency of the interactive behavior generated during user operations such as camera switch or video playing or pausing. This latency affects user experience in real-time operations.

The throughput stability is proposed in this paper, according to the research on the real-time streaming media protocol in ITU-T P.1201.

Table 4-2 Core factors affecting the uplink multimedia message transmission services

Factor	Impact
Transmission waiting time	Under poor network performance, the transmission waiting time is long, adversely affecting user experience. However, in this scenario, the transmission waiting time is closely related to the size of the file to be transmitted. Therefore, it cannot reflect the objective service quality, and is not recommended to be used for evaluation and monitoring.
Uplink throughput	Throughput is the key to guaranteeing the quality of video transmission. Multimedia message transmission is essentially a file uploading process. Throughput assurance is the key. Unlike real-time streaming media services, uploading services are best-effort services, which reflect the maximal uplink transmission performance of pipes.
Throughput swing index	In upload services, the file size varies according to the enterprise requirements. When a large file is transmitted, the transmission latency is high. In this case, the throughput fluctuation affects the waiting time and user experience.

4.1.2 Indicator System

Table 4-3 Indicator system of uplink real-time streaming media transmission services

Layer	Protocol	Indicator	Indicator Measurement Description
Comprehensive Score	E-Score		
Media quality index (MQI)	Media Quality Index		
	MPEG	Video resolution	Generally, the resolution of the camera is fixed.
	MPEG	Video bit rate	Generally, the average bit rate of the camera is fixed.
	MPEG	Video frame rate	Generally, the average frame rate of the camera is fixed.
	MPEG	Video encoding and decoding	H.264/H.265 image compression encoding
	MPEG	Audio bit rate	Generally, the average bit rate of the camera is fixed.
	MPEG	Audio frame rate	Generally, the average frame rate of the camera is fixed.
Interaction quality index (IQI)	Interaction Quality Index		
	SDK	Interactive latency	Latency from the time a control message is sent to the time the

Layer	**Protocol**	**Indicator**	**Indicator Measurement Description**
			message is responded.
	MPEG	Encoding latency	Encoding latency of the camera
	MPEG	Decoding latency	Decoding latency of the decoding server
	RTSP	Video playback latency	Latency from PLAY to 200OK
	RTSP	Video pause latency	Latency from PAUSE to 200OK
	RTCP	Round-trip latency	Calculated based on the RTCP timestamp
	RTCP	Latency jitter	Calculated based on the round-trip latency
Presentation quality index (PQI)	Presentation Quality Index		
	SDK	Slice	Obtaining the decoding server SDK
	SDK	Stall	Obtaining the decoding server SDK
	RTP	Average uplink throughput	Calculating the average value at the flow level
	RTP	Uplink peak throughput	Setting the sampling window and maximum measurement value
	RTP	Uplink throughput swing index	Fluctuation of the experience throughput
	RTP	Average transmit packet discard rate	Calculated based on the RTP sequence
	RTP	Burst transmit packet discard rate	Calculated based on the RTP sequence
	RTCP	Round-trip latency	Calculated based on the RTCP timestamp
	RTCP	Latency jitter	Calculated based on the round-trip latency

In the uplink multimedia message transmission scenario:

Table 4-4 Indicator system of uplink multimedia message services

Category	Indicator
Comprehensive service quality evaluation	Transmission quality index/E-Score
Uplink transmission quality	UL Average Throughput
	UL Peak Throughput
	UL Throughput Swing Index

4.1.3 Modeling Method

Figure 4-1 E-Score modeling framework for uplink real-time streaming media transmission services

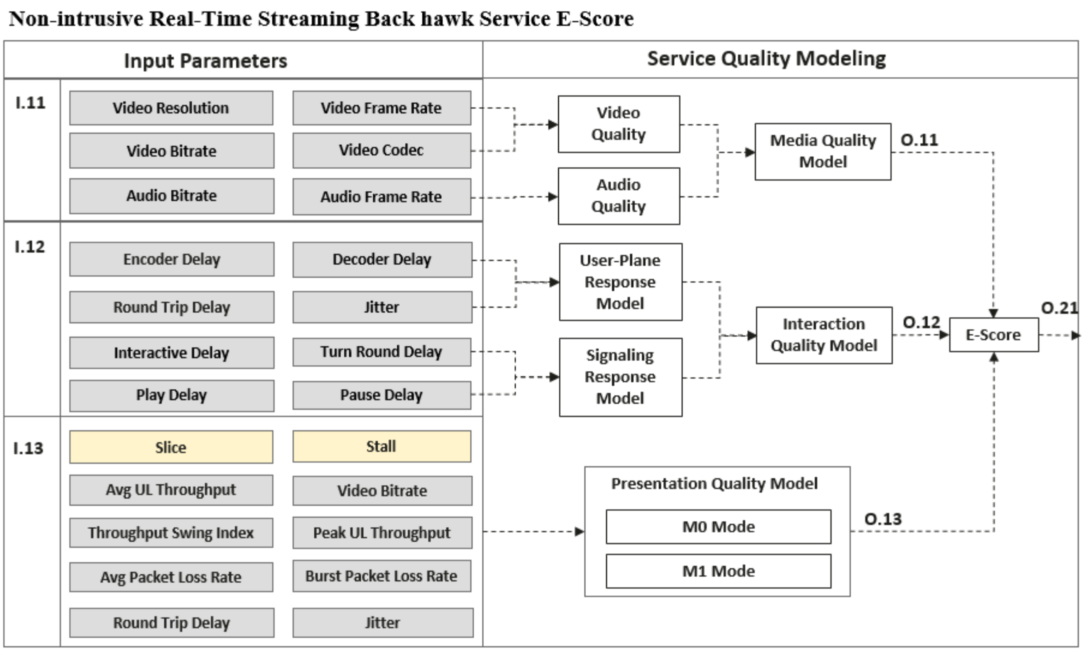


Table 4-5 I.11–input parameters of media quality

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Video bit rate	$VideoBr$	Float, kbps	Per segment	Mode 0	MQI
Video frame rate	$VideoFr$	Integer	Per segment		
Measurement interval	TPD	Float, ms	Per segment		

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Video resolution	Res	Length x width 2880 x 1600	Per segment		
Video encoding	$VideoCodec$	One of: H264-baseline, H264-high, H264-main, H265-high, H265-main	One Session		
Object moving speed	*objSpeed*	Float, Km/s	Per segment		
Object distance	*obDistance*	Float, m	Per segment		
Screen size	*screenSize*	Float, in	One Session		
Audio bit rate	$AudioBr$	Float, kbps	One Session		
Audio encoding	$AudioCodec$	Integer	One Session		

Table 4-6 I.12–Input parameters of interaction quality

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Encoding latency	D_e	Float, ms	Per Segment	Mode 0	IQI
Decoding latency	D_d	Float, ms	Per Segment		
Interactive latency	D_i	Float, ms	Per Segment		
Round-trip latency	D_{rt}	Float, ms	Per Segment	Mode 0 Mode 1	
Latency jitter	$jitter$	Float, ms	Per Segment		
Playback latency	$PlayDelay$	Float, ms	One Session		
Pause latency	$PauseDelay$	Float, ms	One Session		

Table 4-7 I.13–Input parameters of presentation quality

Input Parameter Description	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Pixelization start time	$Slice_{BT}$	Float, ms	Per Segment	Mode 0 Mode 1	PQI
Pixelization end time	$Slice_{ET}$	Float, ms	Per Segment		
Frame freezing start time	$Stall_{BT}$	Float, ms	Per Segment		
Frame freezing end time	$Stall_{ET}$	Float, ms	Per Segment		
Video bit rate	$VideoCodec$	Float, kbps	One Session		
Average throughput	Thr_{avg}	Float, kbps	Per Segment		
Peak throughput	Thr_{max}	Float, kbps	Per Segment		
Throughput upward swing index	Thr_{ulsi}	Float, %	Per Segment		
Throughput downward swing index	Thr_{dlsi}	Float, %	Per Segment		
Average packet loss rate	ppl	Float, ms	Per Segment		
Burst packet loss rate	bpl	Float, %	Per Segment		
Latency jitter	$jitter$	Float, ms	Per Segment		

- The formula for calculating the uplink real-time streaming media service indicators is as follows:

2. Media quality (enhanced based on ITU-T P.1201)

[E-Score]

$$EScore = \omega_1 * MQI + \omega_2 * IQI + \omega_3 * PQI$$

[Media quality index]

$$MQI = c1 * QA + c2 * QV$$

[Audio quality]

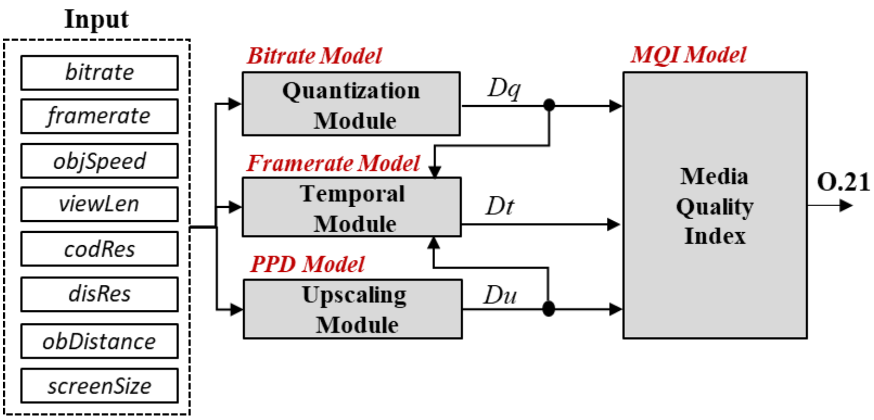
$$QA = I_A - QcodA$$

QA is the predicted audio quality. $QcodA$ is the quality impairment caused by audio compression. I_A is the impact of different audio encoding formats $AudioCodec$. The score of the PCM encoding format is 100, and other audio encoding formats (AAC-LC, HE-AACv2, MPEG1-LII, AC3) all have quality impairment.

$$QcodA = a_{1A} * \exp(a_{2A} * AudioBr) + a_{3A}$$

[Video quality]

Figure 4-2 Modeling framework for video quality index



- Dq (Quantization Degradation): impact of a unit quantity of quantized bits on video quality
- Dt (Temporal Degradation): impact of time complexity on video quality
- Du (Upscaling Degradation): impact of spatial complexity on video quality

$$QV = f\left(D_q, D_u, D_t\right)$$

$$D_q = f_1(disRes, bitrate, sceneType)$$

$$D_u = f_2(disRes, codRes, obDistance, screenSize)$$

$$D_t = f_3(framerate, objSpeed, viewLen, D_q, D_u)$$

[Presentation quality index]

[Mode0]

$$PQI = b_{1V} * \log((b_{2V} * FreezingRatio + b_{3V} * SlicingRatio) + 1)$$

Of which:

$$FreezingRatio = \frac{FreezingDuration}{TotalDownloadDuration}$$

$$SlicingRatio = \frac{SlicingDuration}{TotalDownloadDuration}$$

[Mode1]

$$PQI = c_{1V} * \log(\left(c_{2V} * Q_{bandwidth} + c_{3V} * Q_{packetloss} + c_{4V} * Q_{jitter}\right) + 1)$$

$$Q_{bandwidth} = b_1 * \frac{Thr}{VideoBr} * \exp(b_2 * ULSI + b_3 * DLSI + b_4)$$

$$Q_{packetloss} = p_1 \times \frac{ppl}{\frac{ppl}{BurstR} + bpl} + p_0$$

$$Q_{jitter} = j_1 * \exp(j_2 * Jitter + j_3)$$

In the formula, $BurstR$ indicates the impact factor of burst packet loss on services compared with random packet loss. For live video, the value is fixed.

[Interaction quality index]

$$IQI = 100 - Q_{up} - Q_{sp}$$

[User-plane response model]

$$Q_{up} = a_{1I} * log(a_{2I} * (D_{rt} + D_e + D_d) + a_{3I} * Jitter + a_{4I})$$

[Control-plane response model]

$$Q_{sp} = b_{1I} * log(b_{2I} * (D_i) + b_{3I})$$

In the formula, D_i can be obtained in different methods. Some applications may directly obtain the value from a control-plane interaction message, and others may obtain the value from a protocol. For a play action in the RTSP protocol, $D_i = PlayDelay$, and for a pause action, $D_i = PauseDelay$. Success rate indicators are not closely related to the network. Therefore, you are advised not to include them into the calculation of interaction quality.

- The formula for calculating the uplink multi-media message service is as follows:

For this best-effort multimedia file uploading behavior, consider the uplink bandwidth and its stability.

$$QM = 100 - I_{bandwidth}$$

$$I_{bandwidth} = \begin{cases} 0, & Thr < Thr0 \\ (b_1 \times Thr + b_0) \times \exp(b_2 \times ULSI + b_3 \times DLSI + b_4), & Thr \geq Thr0 \end{cases}$$

NOTE

1. There is no technical difficulty in measuring the throughput and the data is easy to obtain. However, you might run into technical difficulties when trying to measure the throughput fluctuation, as it depends on products or small-granularity statistics for implementation (e.g. probe-based dotting, such as MR data implementation). Additionally, the throughput fluctuation measurement must be performed in a specific time window, because indicator value varies significantly in different time windows.
2. Latency measurement is also technically difficult, as in most scenarios, it is difficult to obtain precise latency metrics. This document provides an end-to-end latency model, which is a non-intrusive small-granularity latency evaluation algorithm. This algorithm can be used to obtain latency indicators, and can break down the latency problem to radio factors such as congestion, interference, and coverage, and provide the requirement boundary of each factor. In this way, the problem can be demarcated for subsequent optimization.

Experience Baseline.

4.1.4 Industrial UAV

UAV is a specific service scenario in 5G use cases. The preceding section mentions the UAV application in the security system. The basic services provided by UAVs are similar regardless of the use cases. This section describes the basic transmission mechanism and basic service scenarios of UAVs.

UAVs mainly use the point-to-point transmission technology, which is classified into the following types:

Service Scenario	Description	Bandwidth	Latency
Real-time flight control command	Commands for flight control. It is an interactive process from sending a request to receiving a response.	10 kbps	20–100 ms
Multimedia data transmission	Photographing and upload of image data, including HD images, infrared images, and VR images.	x–x Mbps	/
Video upload	Different cameras are used in different application scenarios. Videos are provided for survey, viewing, searching, and patrol.	Several to dozens of megabytes	20–100 ms

Network requirements in industrial application scenarios:

Application Scenario	Service	Uplink Rate	Control Latency	Coverage Height	Coverage Range
Logistics	Autonomous flight	200 kbps	< 100 ms	/	Urban, suburban, and rural areas
	Manual takeover based on HD videos	25 Mbps	< 20 ms	< 100 m	
Agriculture	Pesticide spraying	300 kbps	< 100 ms	< 10 m	Rural areas
	Land survey	20 Mbps	< 20 ms	< 200 m	
Patrol Security Rescue	4K video upload	25 Mbps	< 20 ms	< 100 m	Patrol covers infrastructure, and security covers cities.
Surveying and mapping	Laser surveying and mapping	100 Mbps	< 20 ms	< 200 m	Urban and rural areas
Live broadcast	4K video upload	25 Mbps	< 20 ms	< 100 m	Cities and tourist attractions
	8K video upload	100 Mbps	< 20 ms	< 100 m	

Requirements on continuous coverage of UAVs in industrial scenarios:

The operators' networks need to offer continuous coverage to meet the requirements of full-route coverage and uplink bandwidth, if the flight range of UAVs exceeds 5 km and real-time video upload is required.

- Forest or marine area patrol and mapping: The 4G network cannot offer coverage and stable bandwidth assurance in the UAV operating environment (outdoor, 100 to 200 meters above the ground). Therefore, the video recording mode is used for post-analysis.
- Fire fighting: The UAV is about 50 meters high in urban areas with good 4G network coverage. The 4G module of the UAV supports real-time upload of 480p or 720p videos.
- Express delivery: Fixed flight routes. No control during flight and no need for real-time video upload.

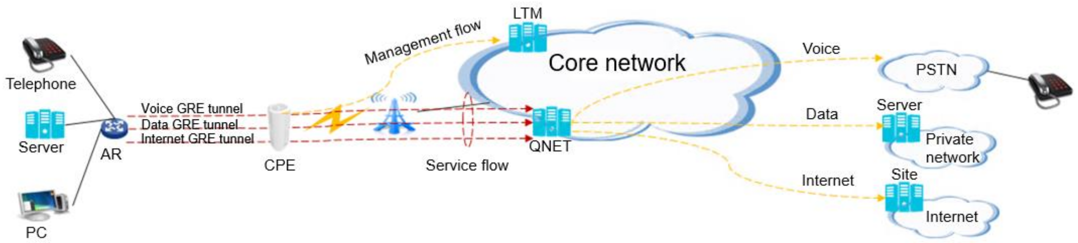
NOTE

5G massive MIMO 3D network coverage and vertical multi-beam features are jointly used for low-altitude coverage. The main lobe covers the ground, and the side lobe covers the low-altitude area for UAVs.

4.1.5 FWA Service

Case source: FWA reconstruction project

Figure 4-3 5G ToB FWA networking



Three Service Scenarios

Access routers (ARs) establish GRE tunnels to differentiate services.

- Voice: enterprise voice services, including IP calls and video conference
- Data: enterprise fax and POS data services
- Internet: enterprise Internet services, including web browsing, video, and mails

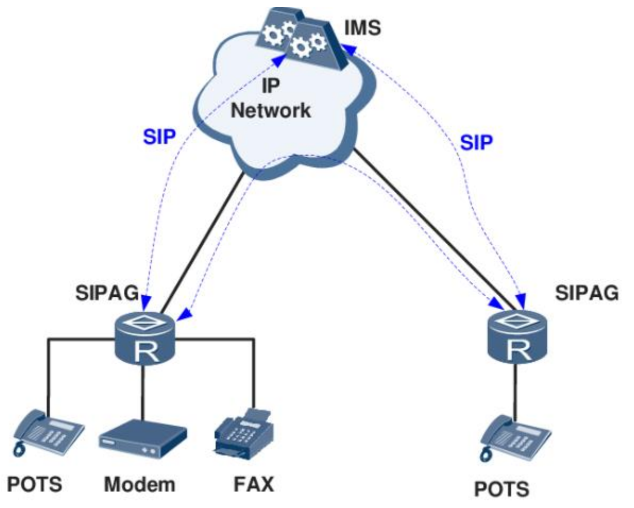
SLA Requirements

- Service availability: transmission channel availability
- Latency and packet loss: transmission channel performance
- Rate: transmission channel performance

Customer requirements:

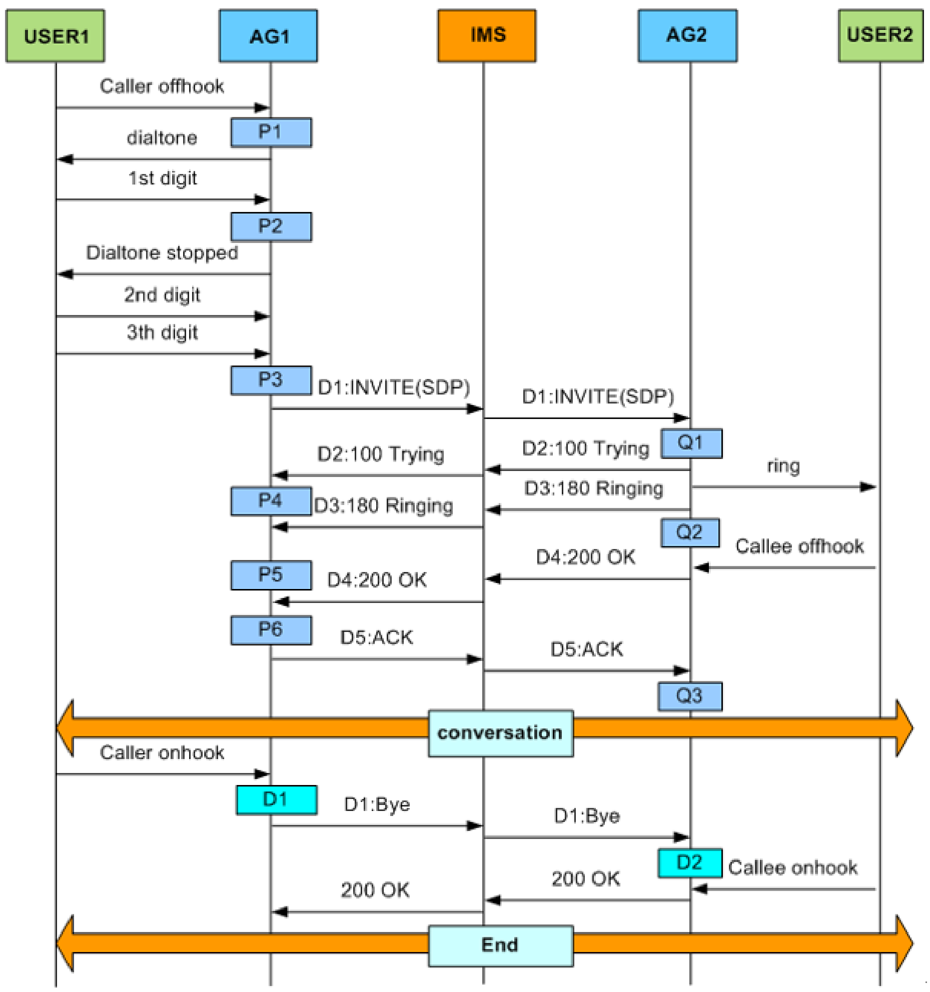
1. FWA transmission performance assurance: not service specific, and involving transmission performance SLA (rate, latency, packet loss, and availability) assurance
2. Service experience assurance: focusing on voice, data, and Internet services

Figure 4-4 Typical networking for enterprise services



Voice services use the Session Initiation Protocol (SIP), and media streams use the Real-Time Transport Protocol (RTP).

Figure 4-5 SIP VoIP service process



4.1.6 Smart City

Widescale use cases in the context of future smart city:

- Enhanced Mobile Broadband (eMBB) services oriented to consumer portable devices, such as smartphones, tablets, and laptops
  - Primarily powered by 4K video and real-time media streams
  - Support for challenging AR and VR applications in local hotspots, if required
- Vehicle-to-infrastructure (V2I) services
  - Entertainment and advertising content for passengers
  - Road and driving conditions and navigation information services (such as parking)
  - Assisted and autonomous driving
- Public utility services
  - Environmental monitoring and intelligent transport system
  - Smart energy, including smart metering and smart grid
- Logistics
  - Sensor data for tracking goods in transit

The following table lists the service requirements.

Item	Impact
Minimum required throughput	This is the minimum throughput that must be guaranteed to ensure user experience. The bandwidth allowed by networks must be above UE-perceived bandwidth.
E2E latency	This is the minimum UE-perceived latency that must be guaranteed to ensure user experience. It includes the round trip time (RTT) as well as the time for receiving and acknowledging data at the application layer, in addition to the time spent by other protocol layers.
Volume per service per day	Average data volume consumed by each device
Number of devices	Expected density of devices for a given service, and can match the traffic density of the given service in an area.
Percentage of scenarios with service coverage	Percentage of devices that are able to receive a given service

The following table lists the network requirements in different service scenarios.

Service Component	Minimum Required Bit Rate	E2E Latency	Reliability and Security	Number of Devices or Users per km^2
eMBB supporting 360-degree video (high throughput but not necessarily low latency)	50 Mbps	< 100 ms	Best effort reliability and Consumer grade security	Up to 150,000
AR/VR-based eMBB with low latency (<10 ms) and high throughput	50 Mbps	< 10 ms	Best effort reliability and Consumer grade security	Up to 150,000
Intelligent traffic signal control (high reliability, low throughput MTC service)	Minimum connectivity	> 100 ms	High reliability and high security	100s of road sensors in the port area.
eMBB service supporting 4k+ video (high throughput MTC service, but not necessarily low latency)	10 Mbps	<100ms	Best effort reliability and high security	10s of video surveillance points in the port area
Low throughput, high density MTC for environmental data analysis or logistics	Minimum connectivity	>100 ms	Best effort reliability and Consumer grade security	10s of thousands of containers in the port area per day 100s of environmental sensors in the port area
eMBB–consumer portable devices (driven by video applications)	10 Mbps (DL/ UL (4k video quality experience)	< 100 ms	Best effort reliability and Consumer grade security	10s of thousands per km^2
V2I–infotainment (eMBB)	10 Mbps DL (4k video quality to at least one passenger)	< 100 ms	Best effort reliability and Consumer grade security	100s of vehicles per km^2
V2I – driver information service (mMTC)	0.5 Mbps DL/UL	< 100 ms	Best effort reliability and Consumer grade security	100s of vehicles per km^2
Environmental monitors, waste management and ITS (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security	100s of devices per km^2

Service Component	Minimum Required Bit Rate	E2E Latency	Reliability and Security	Number of Devices or Users per km^2
Smart meters–sensor data, meter readings, individual device consumption (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security	10s of thousands per km^2
Smart grid sensor data and actuator commands (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security	10s of thousands per km^2
Logistics sensor data for tracking goods (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security	Up to 10k items to track per km^2
V2I–assisted driving	0.5 Mbps DL/UL	< 100 ms	High reliability	100s of vehicles per km^2

4.1.7 Massive Connectivity Services

3GPP cellular systems were primarily designed for human voice and data usage, with less focus on the needs of machines. Backwards compatibility must be considered for all requirements, as both CIoT and IIoT devices will likely become mainstays of the industry and will impact the migration of technology.

1. **Zero Complexity**

One of the more demanding requirements is the lowering of device complexity to virtually zero. This will have a positive impact on the overall cost of devices, as silicon expenses are almost entirely eliminated. In this regard, 3GPP has identified several features that are not required for MTC devices and could significantly reduce device complexity. Most notably, it was proposed that LTE limits device capability to a single receive RF chain, restricting supported peak data rates to the maximum required by IoT applications while also reducing the supported data bandwidth as well as support for half duplex operation. Standardization is required to ensure that maintaining system performance with normal 3GPP devices with additional scheduler restrictions is achievable to serve these low complexity devices. Considering the schedule, 3GPP has closed some of the complexity reduction specifications for LTE in Release 12, with the remaining (as well as all-new) complexity items set to be addressed in Release 13 or later.

2. **Longer Battery Life**

A large percentage of IoT devices rely on battery power, some of which are located in remote areas where replacing or charging batteries may not be possible or economically feasible. Device miniaturization continuously reduces the physical size of batteries, meaning the total available energy in a battery may not increase even though the actual technology is improving. This requires the communications modules of IoT devices to be so energy efficient that decades-long battery life can be ensured. Battery life of 10 years is already feasible for infrequent data transmissions with both LPWA technologies and LTE Release 12. However, one of the challenges for 5G is to achieve the battery life of more than one decade for devices delivering frequent data transmissions.

3. **Extended Coverage**

Many IIoT and even CIoT applications require wide coverage, examples of which include smart metering and factory automation with basement coverage. Many connection business models only work if the majority of devices on the network can be connected. However, due to the nature of radio channels, it is costly to provide 100% coverage in indoor locations such as basements. There is also a need to reach that last few percent of devices positioned in challenging locations without adding significantly to the total cost of the complete solution. Increasing the number of base stations is a possible solution, but comes at the extra cost of site acquisition/rental, backhaul provisioning, and other expenses. A more viable approach is to improve coverage levels in certain critical application contexts without adding significantly to the overall cost of the solution. In this regard, 3GPP is stipulating MTC devices featuring low complexity and improved coverage to facilitate a scalable IoT uptake. Notably, coverage improvement can be achieved by repetition of information, with more details provided in 3GPP Release 12, Stage 3.

4. **MTC User Recognition and Control**

Most cost-efficient MTC devices have integrated SIM cards. However, certain scalability, configuration, and complexity reasons mean that this is not always the case. In these scenarios, access permission regulation is required using previously defined SIM profiles. SIM cards typically contain subscriber IMSIs for directly connecting to an HLR, and these include details about subscribed MTC services and feature profiles. Operators are already able to support customized MTC services based on subscription profile, such as optimal data packet size and optimal routes with dedicated access point name (APN) for MTC services. Through the IMSI, specific charging policies for MTC subscriptions are provisioned by operators, who have complete control over the subscribers allowed on the network. 3GPP is likely to define one or more new LTE UE categories for MTC, and this will become one of the primary methods used to identify and isolate MTC devices if they are impacting performance of the network, as well as restrict access for MTC devices. A common concern among operators is the restriction of access to roaming devices, as operators must be able to identify such roaming MTC devices from an MTC specific UE category and restrict access to the devices if they do not wish to service them.

5G mMTC Customer Requirements

The number of "things" connected through IoT systems today falls short of the forecasts made in previous years, with IoT deployment lagging behind market research predictions. This is a surprising development for many people, as providing the planet with IoT capabilities will result in considerable operational savings and financial gains. In order to properly understand potential market dynamics, the following three concepts for the roll-out of successful new technologies must be considered:

- Provisioning of the technology
- Tested business models that link supply with demand
- Strong market demand

A large number of connection technologies and standards have been tested and are available today, with many already successfully deployed worldwide.

From a business modeling perspective, numerous models are currently available and some have even been successfully tested in actual commercial deployments. For example, in the smart city market, the city hall uses smart parking sensors, smart garbage bin sensors, and smart street-lighting sensors. In addition to guiding drivers to vacant parking spots (and thereby reducing driving time and pollution), the smart parking sensors correlate the occupancy data with the payment data, the latter of which allows infringements to be spotted more efficiently, thereby increasing financial income. Smart bin sensors can detect exactly when the bin needs to be emptied, improving pick-up schedules and saving money. Smart street-lighting sensors regulate lamp usage according to ambient light conditions and movement on the street (if no movement is detected in the early hours of the morning, the lights switch off). This can lead to estimated savings of 30% on the electricity bill.

Given the extensive supply of technology and the strong business models in place, why has IoT still developed slower than expected? The reason lies with market demand, which remains consistently low. This is entirely normal in scenarios involving new technologies and markets. For example, the Internet took more than a decade to gain widespread use, as people had become accustomed to offline accounting and shopping and found it difficult to break these habits. Manual techniques for measuring city air pollution have been in practice for many years, and agencies are reluctant to now switch to autonomous sensors in the context of IoT smart cities. This general hesitancy is arguably the biggest challenge for the IoT industry today, which is to create a genuine demand among industries and consumers. Once demand is generated, and procurement and supply chains are adapted, IoT will embrace fast growth, just as the Internet did all those years ago.

The Internet has undergone a massive transformation from being infrastructure-driven (Ethernet cables, routers, and computers) to business-driven (such as Facebook and Google). Internet of Things is undergoing a similar transformation. From the perspective of the telecommunications ecosystem, the 5G Internet of Things will introduce an evolution of business models.

4.2 Related Technologies

4.2.1 UDP

With the growing trend of strict real-time requirements and high reliability of 5G services, traditional TCP is becoming gradually obsolete, replaced instead by UDP-based protocols.

Table 4-8 Comparison of three live stream transmission protocols

Implementation	Advantage	Disadvantage	Example of Implementation
TCP	• Fairness • Guaranteed delivery	• Possible buffering at receivers • Limited latency control	• HLS • MPEG-DASH
UDP	• High throughput • Low latency	• Lost packets not handled	• RTP • RTMP
ARQ	• High throughput • Guaranteed delivery within window	• Huge latency (but often fixed)	• ZIXI • SRT • RIST • LRT

Reference: Official NAB 2019 Conference Paper: White-paper-Cloud-ingest-of-live-video–An-open-approach-to-RIST-SRT-and-retransmission-protocols

Due in part to the availability of open source options, but more importantly because of retransmission technology being more easily implemented by software, the number of vendors in the Internet contribution space is extensive, with new entrants continuously entering the market. These have been chosen because of their wide applicability, large ecosystems, and openness.

ZIXI, Secure Reliable Transport (SRT) and Reliable Internet Streaming Transport (RIST) are typical examples.

Protocol	SRT	RIST	ZIXI
Firewall Traversal	Yes, both sides	Yes, sender only (both planned)	Yes, both sides
FEC Support	No (planned)	Yes SMPTE 2022-1	Yes, proprietary content awareness
Encryption	Yes AES 128/256	No (Planned)	Yes, AES 128/256 & DTLS
Path Protection	No (SMPTE 2022-7 planned)	Yes (bonding optional)	Yes, SMPTE 2022-7, bonding, primary/standby
Null Packet Compression	Yes	No (Planned)	Yes

4.2.1.2 ZIXI

- Dynamic evaluation of link quality
- Content-aware bandwidth optimization
- Dynamic dejitter buffer and empty bit sequence compression option
- Hybrid intelligent error correction mechanism
- Content-aware forward error correction (CA-FEC)
- Restoration of lost packets for content-aware automatic repeat query (ARQ)
- UDP unicast or multicast seamless bit rate adaptation
- Support for diverse industry-standard video codecs and formats (such as MPEG-TS and RTMP)

  ZIXI provides unprecedented interoperability for more than 100 partners and original equipment manufacturers (OEMs), and over 10,000 real-time video channels in more than 100 countries and regions.

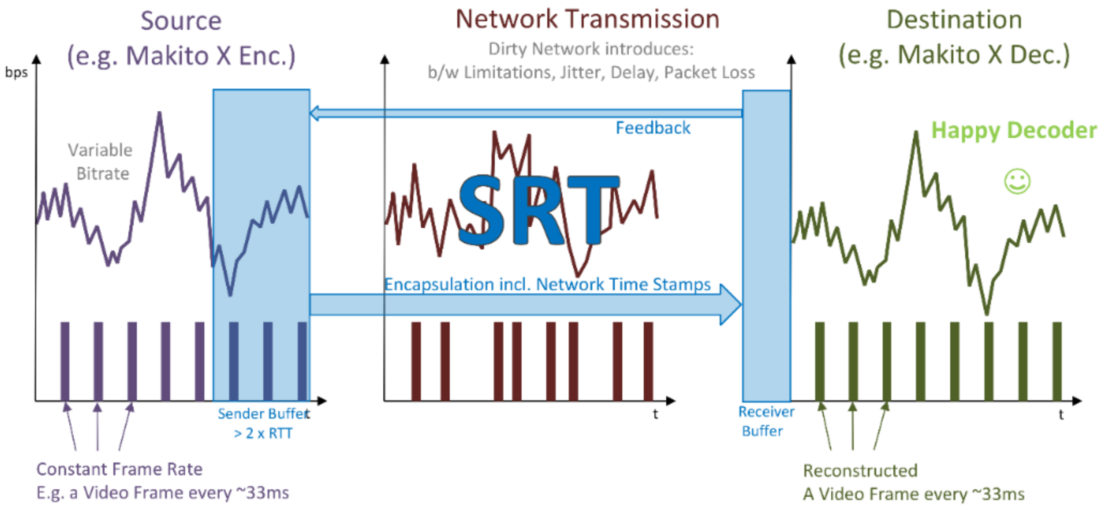
https://zixi.com/

4.2.1.3 SRT

Secure Reliable Transport (SRT) is an open-source commercial toolkit, developed and released by Haivision Systems. It introduces extended and customized functions based on UDP-based data transfer (UDT) protocols to implement packet loss detection, latency control, and video encryption, and enables users to deliver commercial P2P video streams.

- Support ARQ, FEC, and video encryption functions

Figure 4-6 SRT effect



VR platform providers have already adopted the UDP-based SRT protocol to improve cloud VR game performance.

SRT Alliance Deployment Guide, v1.1.

http://www3.haivision.com/srt-alliance-guide

4.2.1.4 RIST

Reliable Internet Stream Transport (RIST): Perhaps the closest to an actual standard is VSFs RIST, or TR-06. The RIST technical recommendation has been agreed upon through a process similar to standardization. It already has more than 30 supporting members, and is growing rapidly.

https://www.rist.tv/members

4.2.1.5 UDT

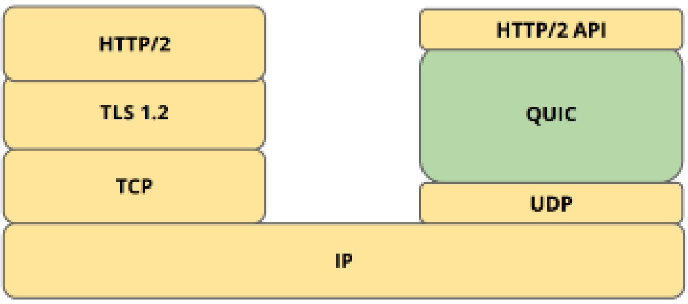
UDP-based Data Transfer Protocol (UDT) is built on UDP and has introduced new mechanisms for congestion and data reliability control. UDT is a connection-oriented two-way application layer protocol and supports both reliable data streaming and partial reliable messaging.

- UDT is duplex, with each entity having two logical parts: a sender and a receiver.
- The sender sends (and retransmits) application data according to flow control and rate control.
- The receiver receives both data packets and control packets and provides feedback.
- The receiver and sender share one UDP port for packet sending and receiving.
- The receiver is responsible for triggering and processing all control events (including congestion control and reliability control) and their related mechanisms, such as RTT estimation, bandwidth estimation, response, and retransmission.
- UDT packs application data into fixed size packets, known as maximum segment size (MSS) packets, unless there is not enough data to be sent. As UDT is intended to transfer bulk data streams, it is assumed that the majority of packets are MSS size. MTU is the optimal value (including packet headers).

- UDT congestion control algorithm combines rate control and window control, with the former tuning packet-sending intervals and the latter specifying the maximum number of acknowledged packets. The parameters used in rate control are updated by bandwidth estimation techniques.

4.2.1.6 QUIC

Quick UDP Internet Connections (QUIC) is a multi-channel concurrent transmission protocol initially designed by Google, with a draft submitted to IETF in 2015. Approximately 7% of global Internet traffic flowed over QUIC in 2018.



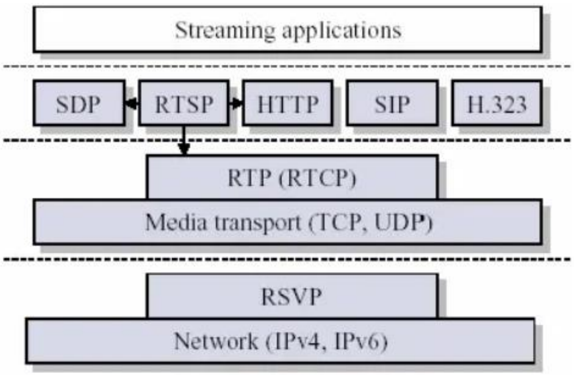
Key advantages of QUIC over TCP+TLS+HTTP/2 include:

- Reduced time for TCP three-way handshake and TLS handshake
- Improved congestion control
- Multiplexing without head-of-line blocking
- Connection migration
- Forward error correction

4.2.2 RTSP

Real-Time Streaming Protocol (RTSP) is an IETF-proposed client-server application-level protocol (in TCP/IP networks) for implementing multimedia playback control. It enables users to control the playback (pause/resume, rewind, and forward) of streaming media as if they were operating a local video player. It is described in RFC 2362.

The following figure shows the RTSP protocol stack.



The RTSP's advantage over HTTP is its support for frame-level control accuracy for video streams, ensuring a high level of real-time performance. As such, RTSP is widely used for H.323 video conference services on enterprise networks and IPTV services on fixed networks. Currently, RTSP is also typically used for ToB video backhaul.

An RTSP message can be a client-to-server request and a server-to-client response.

The following table describes the RTSP methods.

Method	Direction	Requirement	Description
DESCRIBE	Client to server	Recommended	The DESCRIBE method retrieves the description of a presentation or media object and allows the Accept header to specify the description formats that the client understands. The DESCRIBE reply-response pair constitutes the media initialization phase of RTSP.
ANNOUNCE	Client to server Server to client	Optional	When sent from a client to a server, the ANNOUNCE method posts the description of a presentation or media object identified by the request URL to a server. When sent from a server to a client, the ANNOUNCE method updates the session description in real-time. If a new media stream is added to a presentation, the whole presentation description will be sent again, rather than just the additional components, so that components can be deleted.
GET_PARAMETER	Client to server Server to client	Optional	The GET_PARAMETER method retrieves the value of a parameter of a presentation or stream specified in the URL. If no entity body is contained, this method can be used to test client-server connection.
OPTIONS	Client to server Server to client	Required	An OPTIONS request can be issued at any time, for example, if the client is about to attempt a nonstandard request, it does not influence server state.
PAUSE	Client to server	Recommended	The PAUSE request causes the stream delivery to be interrupted temporarily. If the request URL names a stream, only playback and recording of that stream is halted. If the request URL names a presentation or group of streams, delivery of all currently active streams within the presentation or group is halted. After resuming playback or recording, synchronization of the tracks is maintained. Any server resources are kept, though servers can close the session and free resources after being paused for the duration specified with the timeout parameter of the Session header in the SETUP message.

Method	Direction	Requirement	Description
PLAY	Client to server	Required	The PLAY method tells the server to start sending data via the mechanism specified in SETUP. A client MUST NOT issue a PLAY request until any SETUP requests are acknowledged as successful. The PLAY request positions the normal play time to the beginning of the range specified and delivers stream data until the end of the range is reached. PLAY requests can be queued: A server queues PLAY requests to be executed in order.
RECORD	Client to server	Optional	This method initiates recording a range of media data according to the presentation description. The times tamp reflects start and end time. If no time range is given, the start or end time provided in the presentation description are used. If the session has already started, recording is commenced immediately. The server decides whether to store the recorded data under the request-URL or another URL. If the server does not use the request-URL, the response must be 201 (Created) and contain an entity which describes the status of the request and refers to the new resource, and a Location header. A media server supporting recording of live presentations supports the clock range format, and the SMPTE format does not make sense.
REDIRECT	Server to client	Optional	A REDIRECT request informs the client that it must connect to another server location. It contains the mandatory header Location, which indicates that the client must issue requests for that URL. It can contain the Range parameter, which indicates when the redirection takes effect. If the client continues to send or receive media for this URL, the client must issue a TEARDOWN request for the current session and a SETUP for the new session at the designated host.
SETUP	Client to server	Required	The SETUP request for a URL specifies the transport mechanism to be used for the streamed media. A client issues a SETUP request for a stream that is already playing to change transport parameters. If the server does not allow the change, it responds with error "455 Method Not Valid In This State". For the benefit of any intervening firewalls, a client must indicate the transport parameters even if it has no influence over these parameters.

Method	Direction	Requirement	Description
SET_PARAMETER	Client to server Server to client	Optional	This method requests to set the value of a parameter for a presentation or stream specified by the URL. A request only contains a single parameter to allow the client to determine why a particular request failed. If the request contains several parameters, the server acts on the request only if all of the parameters are set successfully. A server must allow a parameter to be set repeatedly to the same value, but it cannot allow changing parameter values. Note: Transport parameters for the media stream must be set with the SETUP command. Restricting setting transport parameters to SETUP is for the benefit of firewalls. The parameters are split in a fine-grained fashion so that there can be more meaningful error indications.
TEARDOWN	Client to server	Required	The TEARDOWN request stops the stream delivery for the given URL, freeing the resources associated with it. If the URL is the presentation URL, any RTSP session identifier associated with the session is no longer valid. Unless all transport parameters are defined by the session description, a SETUP request has to be issued before the session can be played again.

RTSP uses sessions to describe the life cycle of a connection. A session is set up by an RTSP client using the SETUP method for a media stream. During the liveness of the session, the RTSP client uses the PLAY, PAUSE, and RECORD methods to control the play, pause, and playback operations of the media stream, respectively. When the media stream is no longer needed, the RTSP client closes the session through the TEARDOWN method.

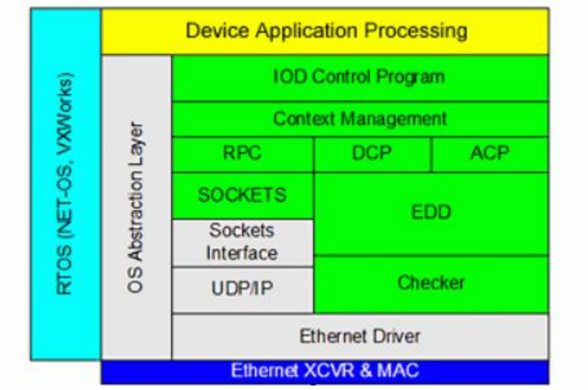
4.2.3 IoT Protocols

4.2.3.1 PROFINET

PROFINET is an open real-time industrial Ethernet communication protocol designed by PROFIBUS & PROFINET International based on TCP/IP and standards for information technology. Since 2003, PROFINET has been part of IEC 61158 and IEC 61784.

PROFINET divides packets into two categories. One is non-real-time packets, which are transmitted based on the TCP/UDP protocol stack and is generally carried out for peer-to-peer communication between a PLC and another PLC or a configuration software unit. The other is real-time packets, known as PROFINET IO, which is transmitted based on Siemens-proprietary underlying protocol stack without relying on TCP/UDP and IP layers to support high-speed I/O data exchange.

Figure 4-7 PROFINET protocol stack



These parts highlighted in green are the protocol stack of PROFINET IO.

Table 4-9 PROFINET packet format

Packet Type	Function	Applicable scenario
PN-PTCP	A Precision Time Control Protocol based on IEEE 1588. Within this context, it is multicast LLDP packets sent by slave nodes at an interval of 200 ms and does not require the master node to reply. PN-PTCP frames are identified by the Ethertype value 0x8892. Four states are defined: Sync, DelayResp, Followup, and DelayReq.	Sent between slave nodes every 200 ms since the start
LLDP	A link layer discovery protocol that enables devices to use TLV elements to send such information as capabilities, management addresses, device identities, and interface identifiers to directly connected neighbors. LLDP frames are identified by the Ethertype value 0x88cc.	Sent between slave nodes every 5s since the start
PN-DCP	A discovery and basic configuration protocol that enables nodes of specified IP addresses to be identified and discovered, as well as IP addresses, gateways, and subnet masks to be configured for them.	It is used only in the PROFINET context.
PNIO_PS	Periodic real-time packets within the PROFINET context between master and slave nodes every 256 cycles (or 8 ms) once a connection is established between them. Sent by both the master and slave nodes, the type of packets are transmitted periodically in real time.	Transmission between master and slave nodes within the PROFINET context in most cases

Packet Type	Function	Applicable scenario
PNIO-CM	Aperiodic packets transmitted between master and slave nodes upon the setup and disconnection between them, which are aperiodically read and written.	Connection setup between master and slave nodes and non-periodic read and write within the PROFINET context

4.2.3.2 ZigBee

The ZigBee protocol is designed to meet the requirements of wireless sensors for low cost, low power consumption, and high fault tolerance. The foundation of Zigbee is IEEE802.15.4. IEEE protocols only process the low layers of the MAC and physical layer, and the Zigbee Alliance has extended the IEEE with network layer protocols and APIs standardized. ZigBee is a new short-distance and low-rate technology that is mainly used for short-range wireless connections. It has its own standards and coordinates communication between thousands of tiny sensors.

Zigbee is a wireless data transmission network platform consisting of up to 65,000 wireless data transmission modules. It is similar to the existing CDMA or GSM mobile communications networks, and each Zigbee network data transmission module is similar to a mobile network base station. These modules can communicate with each other on the network, and the distance between each network node can be extended from the standard 75 meters to hundreds of meters or even several kilometers. In addition, the entire ZigBee network can be connected to other existing networks.

Generally, Zigbee technology can be used for wireless transmission for applications meeting any of the following conditions:

- Transmission involves many outlets that require data to be collected or monitored;
- A small amount of data needs to be transmitted and the device cost must be kept low;
- The reliability and security requirements are high;
- The device uses a battery power supply;
- The device is small and cannot house large rechargeable batteries or power modules;
- The terrain is complex and there are many monitoring points, requiring wide network coverage;
- The device is located in the coverage holes of the existing mobile network;
- Telemetering and remote control system uses the existing mobile network for low data transmission;
- The GPS is costly or ineffective for locating moving objects in some areas.

It should be noted that no specific routing protocol is specified in the published Zigbee V1.0, and the specific protocol is implemented by a protocol stack.

ZigBee is designed for industrial and home automation, as well as telemetering and remote control, for example, for automatic illumination control, wireless data collection and monitoring of sensors in oilfields, electric power, mining, and logistics management.

Features of ZigBee technologies met the following requirements of the industrial field on wireless data transmission:

- Low power consumption, low data volume, and low cost
- Free ISM frequency band (2.4 G)
- Direct sequence spread spectrum (DSSS) with high anti-interference performance
- High confidentiality (64-bit factory number and AES-128 encryption), high integration, and high reliability

The node modules support automatic dynamic networking, adopt the topology structure including the mesh network, and use the collision avoidance mechanism. Information is transmitted in the entire ZigBee network through automatic routing, ensuring reliability.

Technical highlights

1. Low power consumption: In low power consumption standby mode, two AA batteries can support one ZigBee node to work for 6 to 24 months or longer. This is a prominent advantage of ZigBee. In contrast, Bluetooth can work only for weeks at a time, and Wi-Fi can work only for hours at a time under the same condition.
2. Low cost: By greatly simplifying the protocol (less than 1/10 of Bluetooth), the demand for communication controllers are reduced. The predictive calculation is based on the 8051 8-bit microcontroller. The host nodes with full function require 32 KB flash memory, and sub-function nodes only need 4 KB flash memory. In addition, ZigBee is exempted from the protocol patent fee. The price of each chip is around US$ 2.
3. Low rate: ZigBee transmits data at a rate of 20 to 250 kbps. It provides an original data throughput of 250 kbps (2.4 GHz), 40 kbps (915 MHz), and 20 kbps (868 MHz), meeting the requirements for low-rate data transmission.
4. Short distance: The transmission distance ranges from 10 m to 100 m. After the radio frequency (RF) transmit power is increased, the transmission distance can be increased to 1–3 km. This is the distance between adjacent nodes. The transmission distance will be even longer if relaying is achieved through routing and inter-node communication.
5. Short latency: ZigBee is quick to respond, taking only 15 ms to wake from the sleep state, and only 30 ms to connect a node to the network, further reducing power consumption. By contrast, Bluetooth needs 3s to 10s and WiFi needs 3s to perform the same operations.
6. Large capacity: The ZigBee network supports star, tree, and mesh topologies, in which one master node manages up to 254 sub-nodes. In addition, the master node can be managed by the upper-layer network node, which can form a large network with a maximum of 65,000 nodes.
7. High security: ZigBee provides three security levels. The lowest level has no security settings, the middle-level uses an access control list (ACL) to block unauthorized data access, and the highest level uses the advanced encryption standard (AES 128). The three levels can be flexibly adopted to determine the security attributes.
8. Unlicensed frequency bands: ZigBee adopts the DSSS technology on the Industrial, Scientific and Medical (ISM) bands, specifically, 2.4 GHz (global), 915 MHz (US) and 868 MHz (Europe).

ZigBee is mainly applicable to the following scenarios:

- Homes and other buildings: controlling the temperature of air conditioning, automatic illumination, automatic curtain rolling, gas metering, and remote control of household appliances
- Industrial control: automatic control of various monitors and sensors
- Business: smart tags

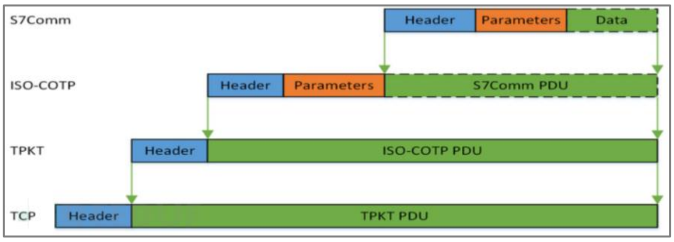
- Public places: smoke detectors
- Agricultural: collection of information about soil and climate
- Medical care: emergency call devices and medical sensors for the elderly and those who lack mobility.

4.2.3.3 S7Comm

The Siemens S7Comm protocol (applicable to S7-300, S7-400, and S7-1200) implements TCP/IP functionality based on the block-oriented ISO transmission service. This protocol is encapsulated in the TPKT and ISO-COTP protocols to enable the protocol data units (PDUs) to be transmitted through the TCP.

It is used for programming PLCs, data exchange between PLCs, access to PLC data from supervisory control and data acquisition (SCADA) systems, and diagnostics.

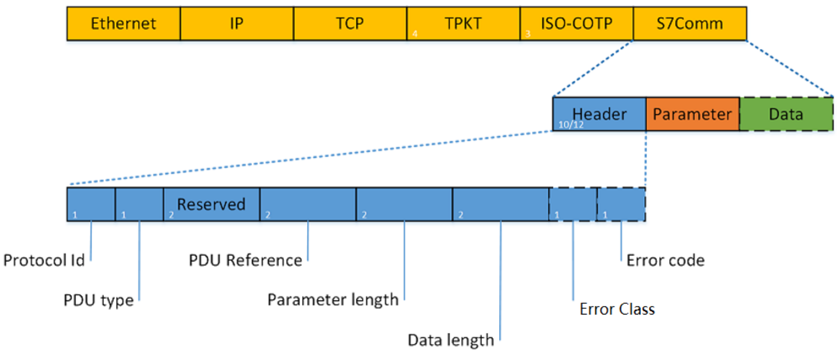
Figure 4-8 OSI model of the S7Comm protocol



The S7Comm protocol consists of three parts:

- Header
- Parameter
- Data

Figure 4-9 S7Comm header structure



PDU type: indicates the type of PDU. The values are as follows:

- 0x01: (JOB, request: job with acknowledgment) A request (for example, to read/write memory, read/write block, start/stop device, or set up communication) sent by the master device;
- 0x02: (ACK, acknowledgement without additional field) A simple acknowledgment without data;

- 0x03: (ACK_DATA, Response: acknowledgment with additional field) Generally a response to a JOB request.
- 0x07: (USERDATA). The parameter field contains the request/response ID (used for operations such as programming/debugging, SZL reading, security function, time setting, and cyclic data).

The following table lists the common function codes used when the PDU type is JOB or ACK_DATA.

Hex	Value
0x00	CPU services
0xf0	Set up communication
0x04	Read Var
0x05	Write Var
0x1a	Request download
0x1b	Download block
0x1c	Download ended
0x1d	Start upload
0x1e	Upload
0x1f	End upload
0x28	PI-Service
0x29	PLC Stop

5 5G ToB Service Characteristic Analysis

5G features large-scale dense deployment of base stations and a small coverage scope, increasing the data rate and supporting ubiquitous access to the network. However, it also imposes new challenges and requirements to the network, such as site selection for micro base stations, energy efficiency, and resource management. To meet the requirements of next-generation mobile services, 5G networks must be deployed such that new base stations are deployed at the minimal CAPEX, particularly in densely populated urban areas where 5G network planning is complex.

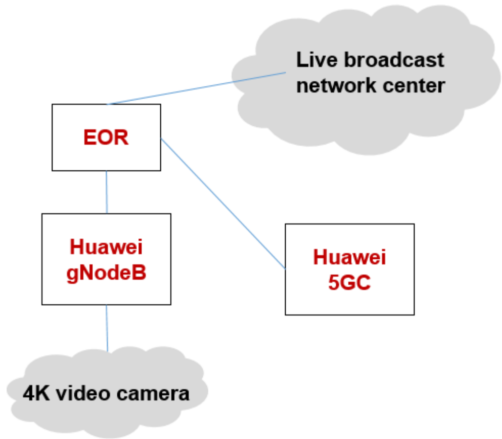
5.1 Multi-Media Transmission

Multi-media transmission includes real-time surveillance video transmission, multimedia message transmission, and UAV video transmission.

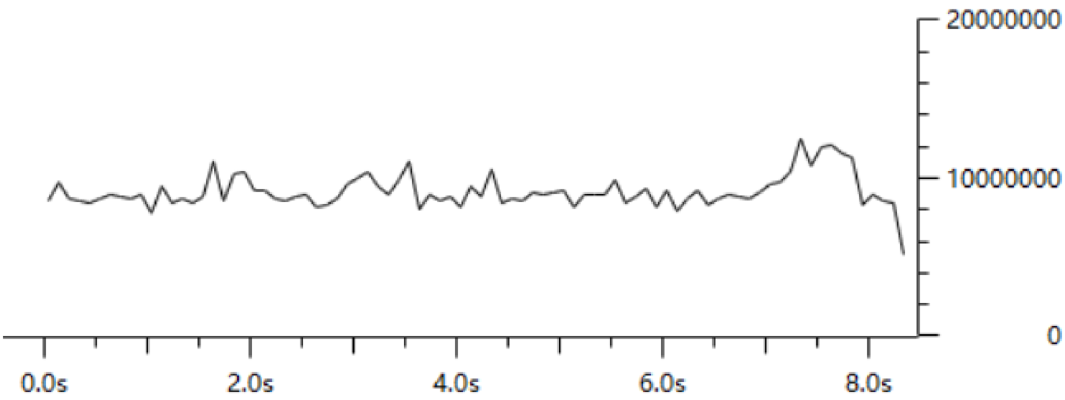
5.1.1 HD Live Broadcast at Site C

[4K live broadcast at site C in China]

Figure 5-1 4K live broadcast networking at site C



The upload service involves real-time uplink data transmission. The uplink transmission rate is approximately 100 Mbps, which is not necessarily the highest capability of a network. The live broadcast service is transmitted based on the real-time encoding rate of the camera, that is, the average bit rate of 4K videos after real-time encoding is 100 Mbps.



According to the analysis of characteristics, the service rate is bottlenecked by insufficient uplink and unstable network bandwidth.

Bit stream analysis:

No.	Time	Source	Destination	Protocol	Length	Identification	Info
22692	5.962101	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7722 (30498)	[MP2T fragment of a reassembled packet]
22693	5.962192	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7723 (30499)	[MP2T fragment of a reassembled packet]
22694	5.962192	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7724 (30500)	[MP2T fragment of a reassembled packet]
22695	5.962193	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7725 (30501)	[MP2T fragment of a reassembled packet]
22696	5.962194	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7726 (30502)	NULL packet
22697	5.962195	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7727 (30503)	[MP2T fragment of a reassembled packet]
22698	5.962195	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7728 (30504)	[MP2T fragment of a reassembled packet]
22699	5.962196	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7729 (30505)	[MP2T fragment of a reassembled packet]
22700	5.962197	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772a (30506)	[MP2T fragment of a reassembled packet]
22701	5.962247	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772b (30507)	[MP2T fragment of a reassembled packet]
22702	5.962249	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772c (30508)	[MP2T fragment of a reassembled packet]
22703	5.962250	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772d (30509)	[MP2T fragment of a reassembled packet]
22704	5.964726	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772e (30510)	[MP2T fragment of a reassembled packet]
22705	5.964726	10.204.32.2	10.114.0.49	MPEG TS	1358	0x772f (30511)	[MP2T fragment of a reassembled packet]
22706	5.964728	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7730 (30512)	NULL packet
22707	5.964729	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7731 (30513)	[MP2T fragment of a reassembled packet]
22708	5.964781	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7732 (30514)	[MP2T fragment of a reassembled packet]

According to the original packet analysis, the 4K live broadcast service has the following six typical features:

1. The data flow is only upstream. (The entire service flow does not have a downstream packet.)
2. All packets are carried over UDP.
3. The length of all packets is 1358 bytes.
4. Wireshark can identify three types of protocols: MPEG TS, MPEG-I, and MPEG PES. (More than 99.99% of packets are MPEG TS packets.)
5. The IP Identifications of almost all packets are consecutive (the value reaches the maximum 65535 and then restarts from 0). There are few cases of inconsecutive IP Identifications.
6. The interval between two packets is much shorter than 1 ms, the average interval being 0.000270 ms). This also complies with the service features of the 4K live broadcast.

- Irregular packet behavior 1: IP Identifications of some adjacent packets are consecutive but the interval is prolonged.

No.	Time	Source	Destination	Protocol	Length	Identification	Info
22731	5.967298	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7749 (30537)	[MP2T fragment of a reassembled packet]
22732	5.967298	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774a (30538)	[MP2T fragment of a reassembled packet]
22733	5.967298	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774b (30539)	[MP2T fragment of a reassembled packet]
22734	5.967299	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774c (30540)	[MP2T fragment of a reassembled packet]
22735	5.967733	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774d (30541)	[MP2T fragment of a reassembled packet]
22736	6.293440	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774e (30542)	[MP2T fragment of a reassembled packet]
22737	6.293446	10.204.32.2	10.114.0.49	MPEG TS	1358	0x774f (30543)	[MP2T fragment of a reassembled packet]
22738	6.293452	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7750 (30544)	[MP2T fragment of a reassembled packet]
22739	6.293453	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7751 (30545)	[MP2T fragment of a reassembled packet]
22740	6.293454	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7752 (30546)	[MP2T fragment of a reassembled packet]

As shown in the preceding figure, the interval between packets 22735 and 22736 is 325 ms, which is far greater than the average value 0.000270s. In addition, the IP Identifications of the two packets are consecutive. There is a high probability that no packet loss occurs on the network between the two packets and no packet is captured. Only an extra latency is introduced. In the entire service flow, there are 10 occurrences where the IP Identifications of adjacent packets are consecutive but the interval between them is long. These occurrences will not be elaborated on in this document.

No.	Time	Source	Destinatio	Protoco	Lengt	IPID	Packet Interval	IP ID Difference
107278	29.430586	10.204.32.2	10.114.0.49	MPEG TS	1358	52385	0.459527	1
66818	18.401589	10.204.32.2	10.114.0.49	MPEG TS	1358	10590	0.434810	1
104288	28.353861	10.204.32.2	10.114.0.49	MPEG TS	1358	48594	0.385808	1
189113	51.321481	10.204.32.2	10.114.0.49	MPEG TS	1358	4904	0.347195	1
22736	6.29344	10.204.32.2	10.114.0.49	MPEG TS	1358	30542	0.325707	1
144454	39.285049	10.204.32.2	10.114.0.49	MPEG TS	1358	24850	0.313194	1
185429	50.278367	10.204.32.2	10.114.0.49	MPEG TS	1358	1117	0.302819	1
147675	40.272301	10.204.32.2	10.114.0.49	MPEG TS	1358	28627	0.300078	1
25789	7.261629	10.204.32.2	10.114.0.49	MPEG TS	1358	34280	0.297445	1
63386	17.205598	10.204.32.2	10.114.0.49	MPEG TS	1358	6800	0.238225	1
106451	28.80706	10.204.32.2	10.114.0.49	MPEG TS	1358	51558	0.044192	1
205329	55.388888	10.204.32.2	10.114.0.49	MPEG TS	1358	21634	0.016707	1
122037	33.090958	10.204.32.2	10.114.0.49	MPEG TS	1358	2443	0.013780	1
41288	11.172392	10.204.32.2	10.114.0.49	MPEG TS	1358	50248	0.012393	1
146471	39.65716	10.204.32.2	10.114.0.49	MPEG TS	1358	27425	0.012128	1
208289	56.16684	10.204.32.2	10.114.0.49	MPEG TS	1358	24592	0.011413	1

According to the statistics, 200 ms is a notable dividing line. Within 58 seconds, the packet interval is greater than 200 ms at 10 points, with the maximum packet interval reaching 459 ms. Based on the characteristics of 4K live broadcast streams, there is a high probability that prolonged intervals cause pixelization of video images.

- Irregular packet behavior 2: IP Identifications of adjacent packets are inconsecutive at certain points in time.

ip.addr==10.114.0.49 && udp.port==5001 && ip.addr==10.204.32.2 && udp.port==1280

No.	Time	Source	Destination	Protocol	Length	Identification	Info
23291	6.302543	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7a3c (31292)	[MP2T fragment of a reassembled packet]
23292	6.302544	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7a3d (31293)	[MP2T fragment of a reassembled packet]
23293	6.302545	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7a3e (31294)	[MP2T fragment of a reassembled packet]
23294	6.302546	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7a3f (31295)	[MP2T fragment of a reassembled packet]
23295	6.302547	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7a40 (31296)	NULL packet
23296	6.309228	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7c2d (31789)	[MP2T fragment of a reassembled packet]
23297	6.309230	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7c2e (31790)	[MP2T fragment of a reassembled packet]
23298	6.309768	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7c2f (31791)	[MP2T fragment of a reassembled packet]
23299	6.312540	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7c30 (31792)	[MP2T fragment of a reassembled packet]
23300	6.312541	10.204.32.2	10.114.0.49	MPEG TS	1358	0x7c31 (31793)	[MP2T fragment of a reassembled packet]

As shown in the preceding figure, the interval between packets 23295 and 23296 is around 6.7 ms, and the difference between IP Identifications is 493. In addition, the sequence numbers of other packets before and after the two packets increase in ascending order by 1. In the entire service flow, there are 35 inconsecutive IP Identifications.

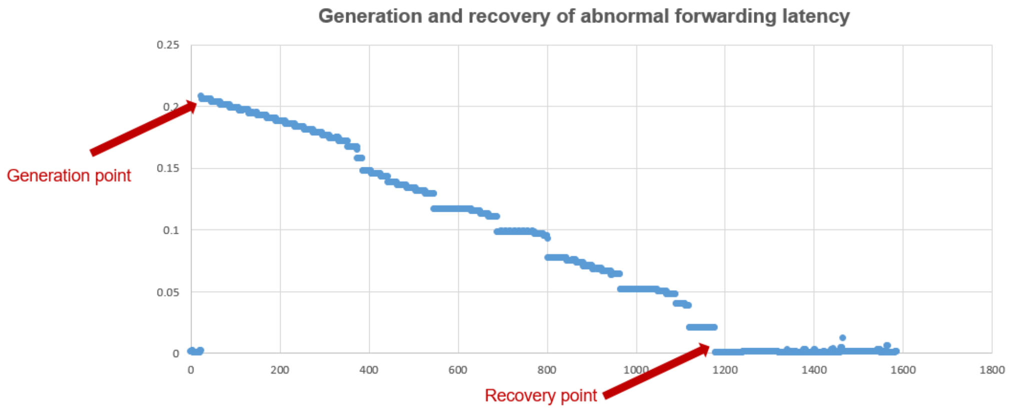
1	No.	Time	Source	Destination	Protocol	Length	IPID	Packet Interval	IP ID Difference
23106	23296	6.309228	10.204.32.2	10.114.0.49	MPEG TS	1358	31789	0.006681	493
23291	104923	28.365712	10.204.32.2	10.114.0.49	MPEG TS	1358	50032	0.000001	362
26129	144976	39.295332	10.204.32.2	10.114.0.49	MPEG TS	1358	25806	0.003733	353
26314	108172	29.442863	10.204.32.2	10.114.0.49	MPEG TS	1358	54122	0.000001	323
26332	67907	18.413642	10.204.32.2	10.114.0.49	MPEG TS	1358	11995	0.000000	317
26460	63905	17.214972	10.204.32.2	10.114.0.49	MPEG TS	1358	7679	0.003116	283
26499	26506	7.277012	10.204.32.2	10.114.0.49	MPEG TS	1358	35472	0.003416	268
63697	104715	28.362026	10.204.32.2	10.114.0.49	MPEG TS	1358	49269	0.002518	249
63882	67913	18.413785	10.204.32.2	10.114.0.49	MPEG TS	1358	12247	0.000001	247
67882	189942	51.33354	10.204.32.2	10.114.0.49	MPEG TS	1358	6255	0.000000	221
67888	26321	7.27199	10.204.32.2	10.114.0.49	MPEG TS	1358	35015	0.002838	201

In the preceding table, the largest IP Identification difference between adjacent packets is 493, and the smallest difference is 2. The interval between adjacent packets with discontinuous IP Identifications is irregular. Specifically, the maximum interval is 6.7 ms, and the minimum interval is 0 ms.

- CN forwarding latency analysis

No.	Time	Source	Destination	Protocol	Length	GTP IP ID	Inner IP ID	No.	Time	Inner IP ID	flg	UGW Forward Latency
66045	5.807495	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58891	29334	66067	5.807922	29334	0	0.000427
66046	5.807496	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58892	29335	66068	5.807922	29335	0	0.000426
66047	5.807604	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58893	29336	66069	5.807923	29336	0	0.000319
66048	5.807604	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58894	29337	66070	5.807923	29337	0	0.000319
66049	5.807605	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58895	29338	66071	5.807923	29338	0	0.000318
66050	5.807605	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58896	29339	66072	5.807924	29339	0	0.000319
66051	5.807605	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58897	29340	66073	5.807924	29340	0	0.000319
66052	5.807606	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58898	29341	66074	5.807924	29341	0	0.000318
66053	5.807606	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58899	29342	66075	5.809895	29342	0	0.002289
66054	5.807607	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58900	29343	66076	5.809896	29343	0	0.002289
66055	5.807607	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58901	29344	66078	5.809896	29344	0	0.002289
66056	5.807607	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58902	29345	67231	6.015565	29345	0	0.207958
66077	5.809896	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58903	29346	67232	6.015565	29346	0	0.205669
66079	5.809897	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58904	29347	67233	6.015565	29347	0	0.205668
66080	5.809897	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58905	29348	67234	6.015676	29348	0	0.205779
66081	5.809897	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58906	29349	67235	6.015676	29349	0	0.205779
66082	5.809898	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58907	29350	67236	6.015677	29350	0	0.205779
66083	5.809992	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58908	29351	67237	6.015677	29351	0	0.205685
66084	5.809992	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58909	29352	67238	6.015677	29352	0	0.205685
66085	5.809993	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58910	29353	67239	6.015677	29353	0	0.205684
66086	5.809993	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58911	29354	67240	6.015677	29354	0	0.205684
66087	5.809993	10.204.32.1	10.114.0.49	GTP <MPEG TS>	1394	58912	29355	67241	6.015678	29355	0	0.205685

1. The most common duration of forwarding latency in the core network is in the milliseconds or even less in some cases.
2. The core network introduces extra forwarding latency during the forwarding process.
3. The abnormal forwarding latency is generated suddenly (before a latency of more than 200 ms is generated, the forwarding latency is at millisecond level). After the abnormal forwarding latency is generated, the forwarding latency gradually returns to normal. The following figure shows the generation and recovery of an abnormal forwarding latency. The Y coordinate indicates the packet forwarding latency, and the X coordinate indicates the packet number. After around 1150 packets are forwarded, the forwarding latency becomes normal.



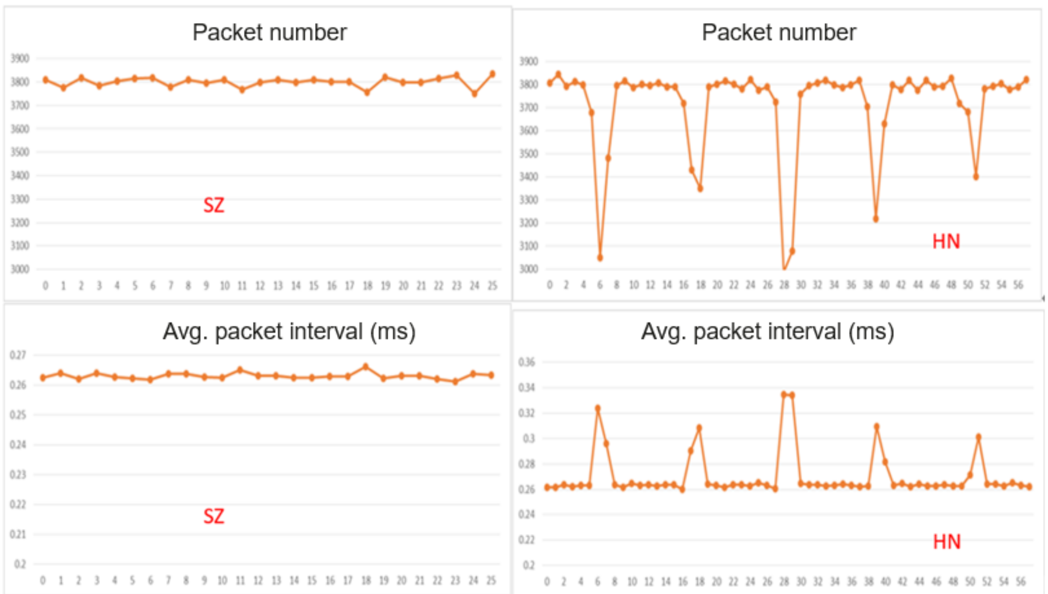
Note: Before the abnormal forwarding latency occurs, the forwarding latency remains at the normal level (ms-level). To facilitate display of the generation and recovery process, the normal forwarding process is not displayed in this figure.

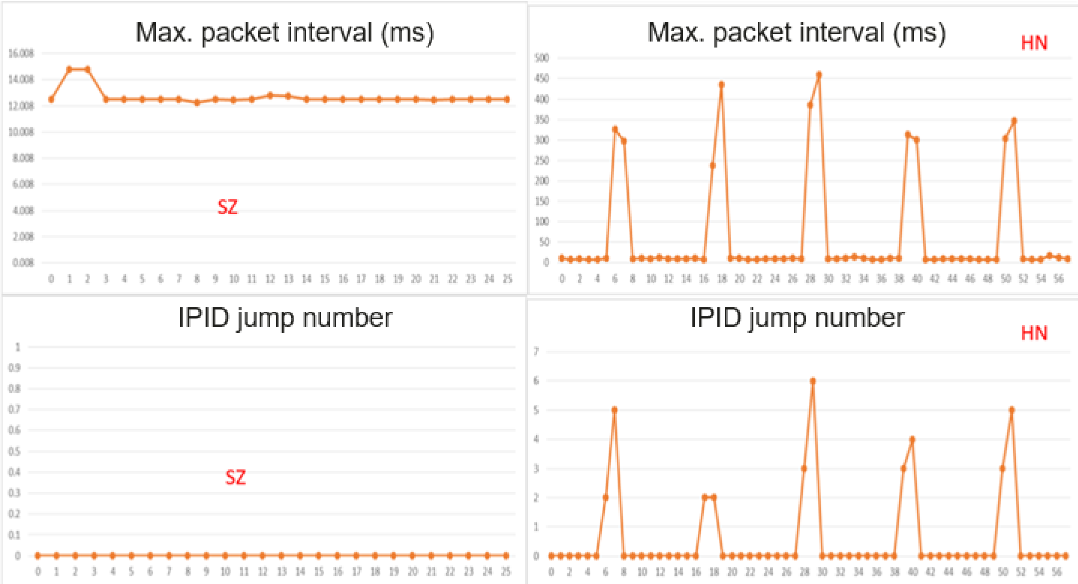
Typical characteristics of the 4K live video broadcast service flow:

1. UDP-based bearer
2. Unidirectional uplink data flow
3. Consecutive IP Identifications of packets
4. Millisecond-level interval between packets

Based on the analysis of pixelization during 4K HD live broadcast, the following indicator system can be constructed to measure the 4K live broadcast service quality:

1. Number of packets/traffic per unit time
2. Average and maximum interval between adjacent packets in a unit time
3. Number of IP ID changes of adjacent packets per unit time (maximum change value)

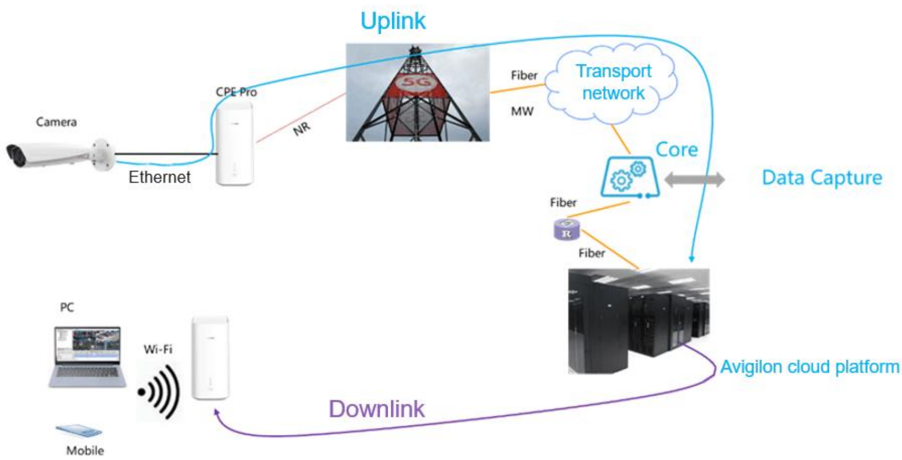




5.1.2 HD Live Broadcast at Site K

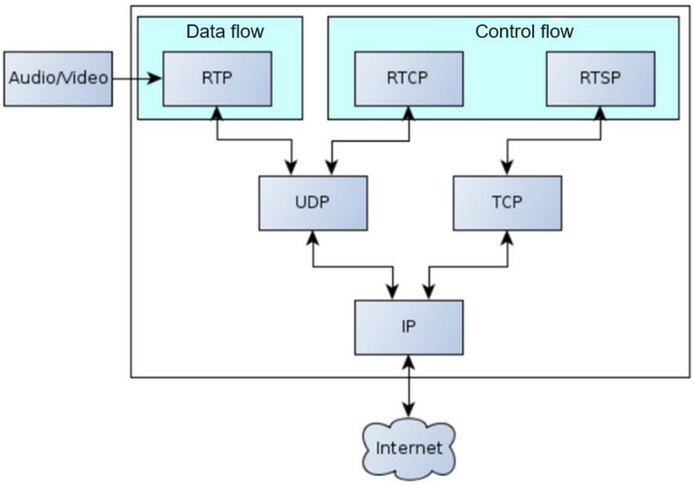
The following figure shows the networking diagram of the live broadcast service at site K.

Figure 5-2 Live video networking at site K



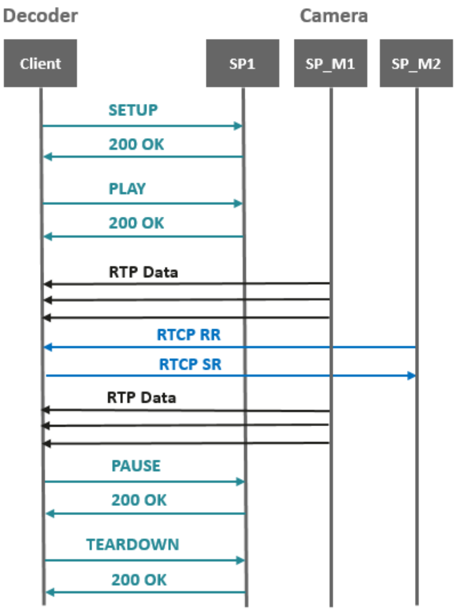
Protocol analysis: Standard RTSP/RTP/RTCP packets are used for transmission. The typical protocol stack for video upload is shown in the following figure.

Figure 5-3 RTSP/RTP/RTCP protocol stack



The video upload process is as follows:

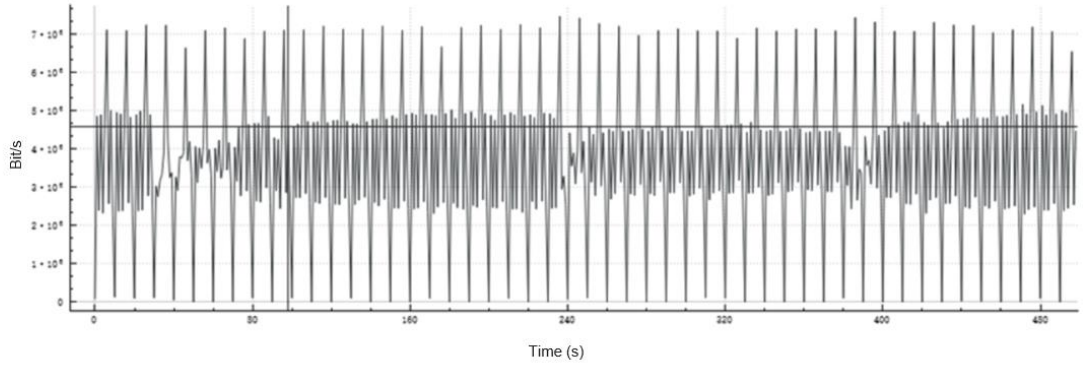
Figure 5-4 RTSP live video message process



RTSP is based on TCP.

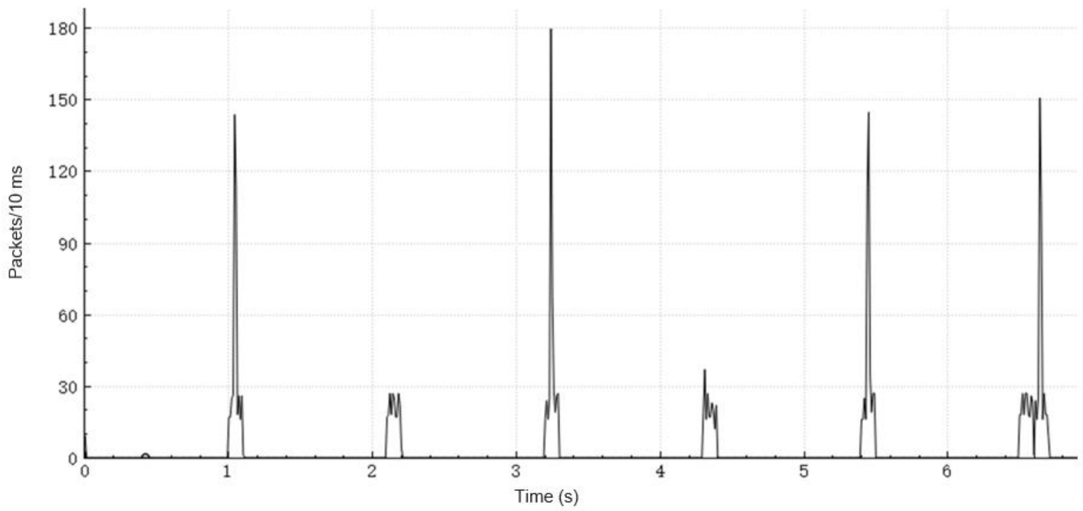
6… 72.269000	10.79.215.139	10.92.63.253	TCP	66	55820 → 554 [SYN, ECN, CWR] Seq=0 Win=8192 Len=0 MSS=1460 WS=256 SACK_PERM=1
6… 72.270000	10.92.63.253	10.79.215.139	TCP	66	554 → 55820 [SYN, ACK] Seq=0 Ack=1 Win=29200 Len=0 MSS=1416 SACK_PERM=1 WS=128
6… 72.270000	10.79.215.139	10.92.63.253	TCP	54	55820 → 554 [ACK] Seq=1 Ack=1 Win=131584 Len=0
6… 72.270000	10.79.215.139	10.92.63.253	RTSP	223	OPTIONS rtsp://10.203.255.253:554/LiveMedia/ch1/Media1 RTSP/1.0
6… 72.271000	10.92.63.253	10.79.215.139	TCP	54	554 → 55820 [ACK] Seq=1 Ack=170 Win=30336 Len=0
6… 72.272000	10.92.63.253	10.79.215.139	RTSP	153	Reply: RTSP/1.0 200 OK
6… 72.272000	10.79.215.139	10.92.63.253	RTSP	249	DESCRIBE rtsp://10.203.255.253:554/LiveMedia/ch1/Media1 RTSP/1.0
6… 72.275000	10.92.63.253	10.79.215.139	RTSP/SDP	694	Reply: RTSP/1.0 200 OK
6… 72.275000	10.79.215.139	10.92.63.253	RTSP	283	SETUP rtsp://10.203.255.253:554/LiveMedia/ch1/Media1/trackID=1 RTSP/1.0
6… 72.281000	10.92.63.253	10.79.215.139	TCP	54	554 → 55820 [ACK] Seq=740 Ack=594 Win=32512 Len=0
6… 72.291000	10.92.63.253	10.79.215.139	RTSP	188	Reply: RTSP/1.0 200 OK
6… 72.291000	10.79.215.139	10.92.63.253	RTSP	256	PLAY rtsp://10.203.255.253:554/LiveMedia/ch1/Media1 RTSP/1.0
6… 72.292000	10.92.63.253	10.79.215.139	TCP	54	554 → 55820 [ACK] Seq=874 Ack=796 Win=33536 Len=0
6… 72.293000	10.92.63.253	10.79.215.139	RTSP	154	Reply: RTSP/1.0 200 OK
6… 72.298000	10.79.215.139	10.92.63.253	TCP	54	55820 → 554 [ACK] Seq=796 Ack=974 Win=130560 Len=0

Data I/O diagram during video transmission:



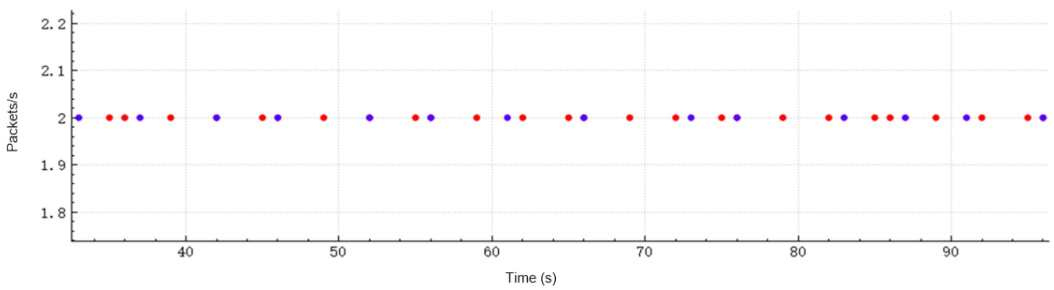
The IP uplink transmission rate is 3 Mbps to 4 Mbps.

Micro I/O graph analysis:



A batch of data is sent per second, and is proactively controlled by the camera video sampling rate. Therefore, the average video upload rate depends on the average video bit rate.

RTCP packets are sent bidirectionally at an interval of around 5s.



RTSP is a control-plane process for live broadcast services. The following indicators are measured:

- Session setup success rate
- Session setup latency
- Video play success rate
- Video play latency
- Video stop success rate
- Video stop latency
- Uplink RTT (TCP bearer)
- Downlink RTT (TCP bearer)

NOTE

The control-plane success rate is very high. Therefore, it is unnecessary to monitor control-plane KPIs as failures are usually caused by problems at the application and content layers, not the network layer. If RTSP is based on TCP, the upstream and downstream RTTs can be monitored to reflect the network-side transmission latency of data packets on the control plane.

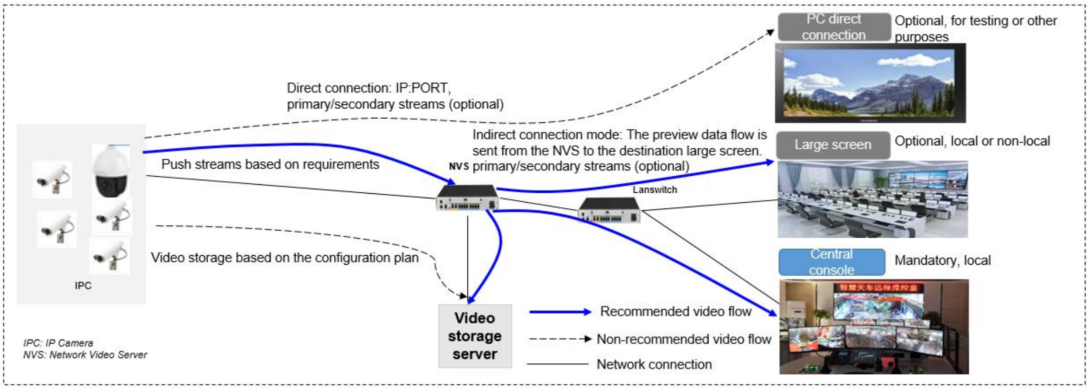
RTP is a user-plane transmission protocol, whereas RTCP is a transmission control protocol. The following counters can be measured:

- Average uplink rate
- Peak uplink rate
- Average packet loss rate
- Burst packet loss rate
- Round-trip latency
- Jitter

5.1.3 Video Surveillance at Site X

The following figure shows the logical network of the video surveillance service.

Figure 5-5 Logical networking of the video surveillance service



Video surveillance data streams:

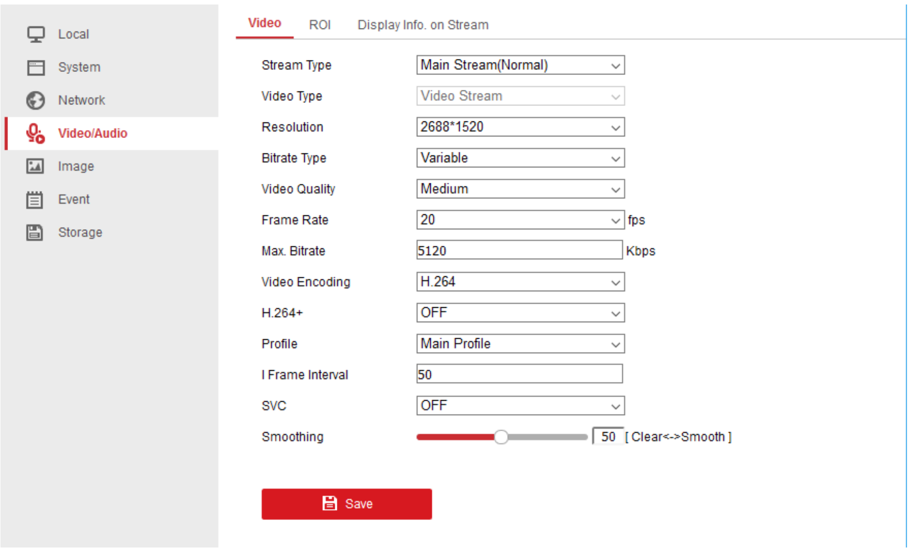
- The camera can push streams to multiple servers simultaneously.

- According to the video surveillance UC design requirements, cameras only need to push streams to only the NVS video server. Directly connecting other servers to cameras is not recommended, as it may increase both the number of video streams and network load.
- The camera pushes streams to the NVS, which then forwards the streams to the storage server, large-screen surveillance system, and central control console.

Primary/secondary streams on cameras:

- The camera supports primary and secondary video stream specifications, meeting video transmission requirements of different quality and bandwidth requirements. It is recommended that primary streams be used for local surveillance and storage. Secondary streams are optional and applicable to low-bandwidth long-distance communication, with seamless connection being the main priority.

Figure 5-6 Configuration description of video surveillance cameras

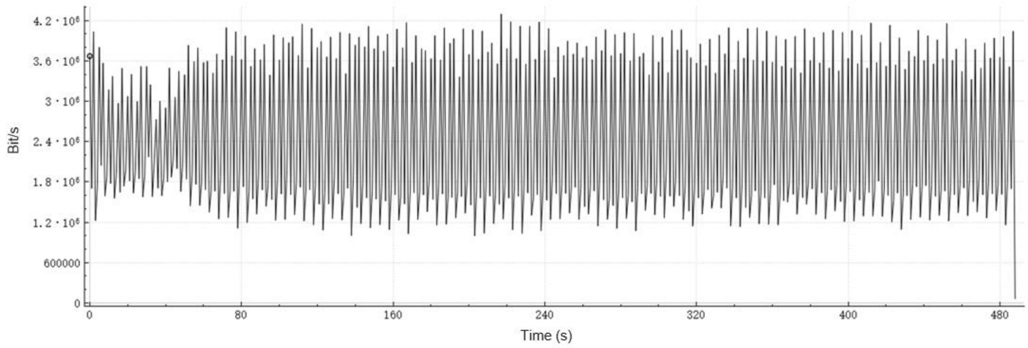


Video camera configuration:

- The camera configuration has a significant impact on the traffic characteristics and performance load of the network. Therefore, you are advised to use the recommended configuration.
- Key configurations: resolution, bit rate type, video frame rate, upper limit of the bit rate, encoding mode, and I-frame interval (1 to 250, 50 by default)

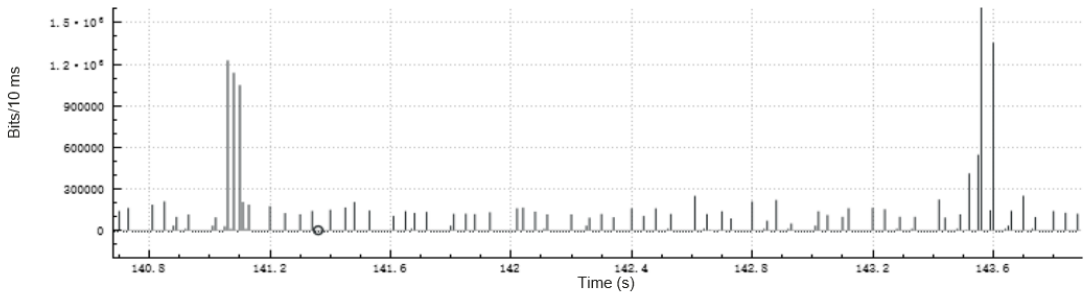
Data is analyzed based on captured packets:

Panoramic video upload, with a backhaul rate of 2.5 Mbps. The following figure shows an I/O graph with a granularity of 1s.



According to the 1-second transmission characteristics, the average rate in a period of time is stable. In particular, the video bit rate is slightly higher than the configured bit rate.

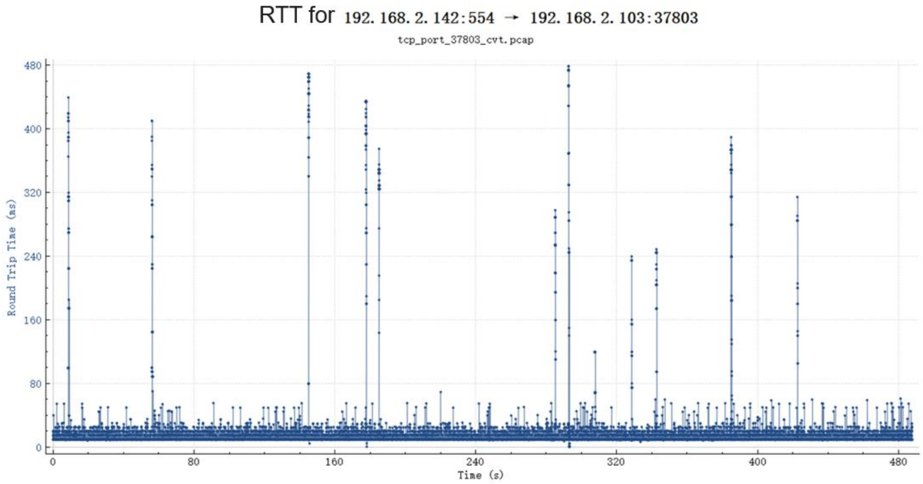
The following figure shows the I/O graph with a granularity of 10 ms.



The transmission behavior of I-frames and P-frames can be observed from a micro perspective (10 ms granularity). The burst peak in the figure is I-frame transmission, and that between two I-frames is P-frame transmission. The transmission period is related to the frame rate and I-frame interval, for example, if the frame rate is 25 and the I-frame interval is 50 (indicating that there are 50 frames in a group of pictures), then an I-frame transmission period is 2s (50/25 = 2), and the average frame interval is 40 ms (1000/25 = 40).

The size of an I-frame is related to the bit rate and resolution. From 2K to 4K resolution, the corresponding bit rate ranges from 2 Mbps to 16 Mbps, and the size of an I-frame is about 500 kbps to 14 Mbps. When an I-frame is sent, the bandwidth requirement is huge. Assuming that the I-frame needs to be sent within 40 ms without affecting the next frame, a corresponding required rate range of the I-frame may be 12.5 Mbps to 350 Mbps. In this case, the actual peak bandwidth required by a single camera is considerably higher than the average bit rate.

The average RTT is about 20 ms; however, it fluctuates greatly. Excessively large RTTs indicate that the network fluctuates significantly.



The following indicators are measured:

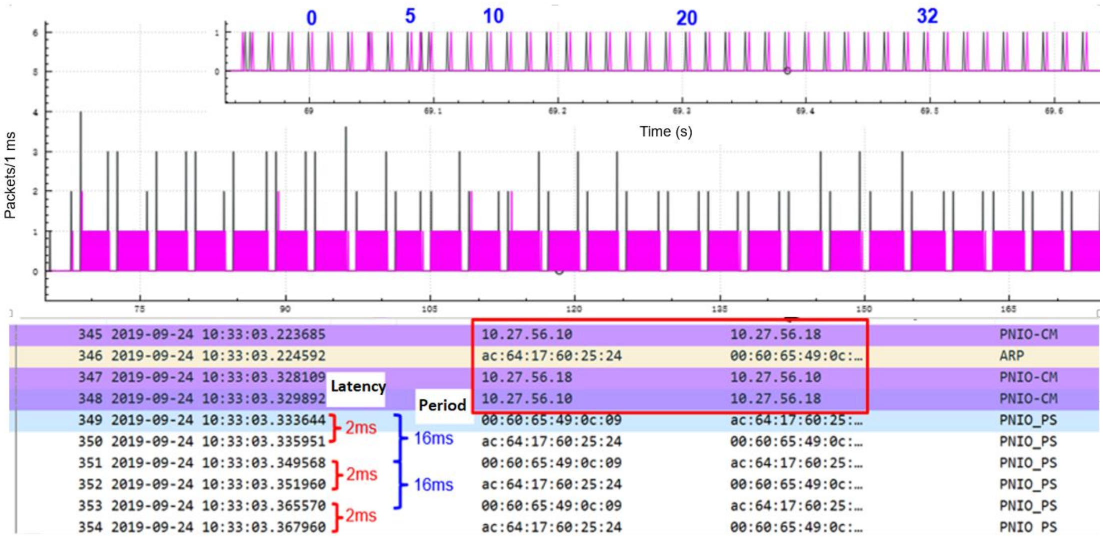
- Average uplink rate
- Peak uplink rate
- Uplink rate fluctuation index
- Uplink RTT
- Uplink RTT poor quality proportion
- Uplink RTT jitter

5.2 Interactive Service Behavior

Interactive service behavior is related to message interaction, specifically, periodic small-packet interaction transmission. PLC is used to implement machine control, and part of this is performed through interaction behavior. Currently, the PLC module of the 5G ToB campus uses the S7Comm and PROFINET protocols. The following describes the transmission features of the two protocols.

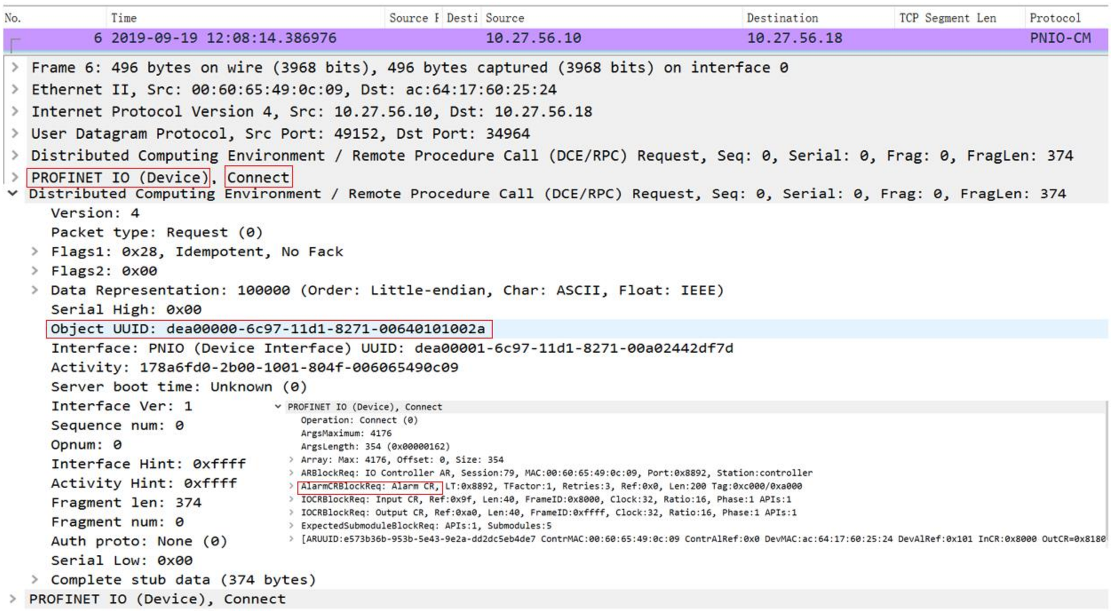
5.2.1 PLC-PNIO

The service characteristics of remote gantry crane in smart ports are analyzed as follows:



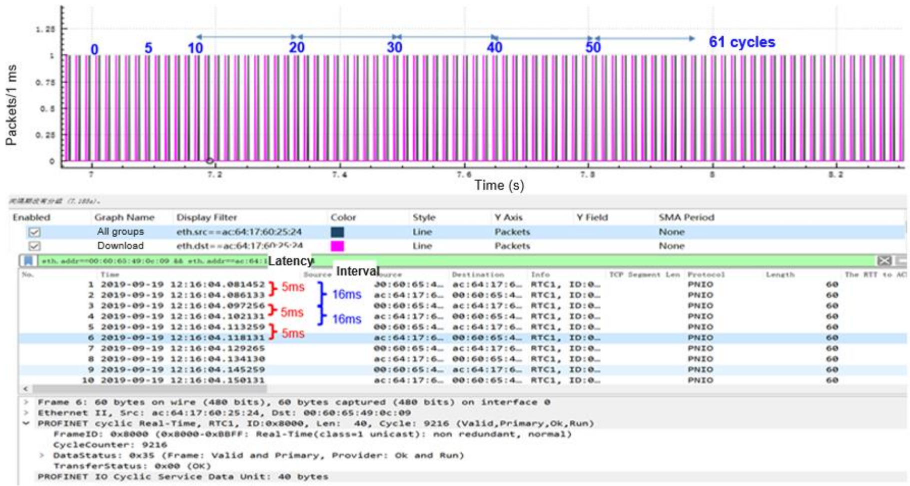
Two types of feature behaviors can be seen:

- 16-ms periodic small-packet interaction: The interaction duration is about 2 ms. The packet size is 60 bytes.
- Uplink burst small packet transmission: The device sends an alarm to the server. The packet size is 500 bytes.



- The device ID (UUID) and alarm request information (AlarmCRBlockReq) are displayed in the PNIO protocol logs.

Behavior analysis in poor-quality scenarios:

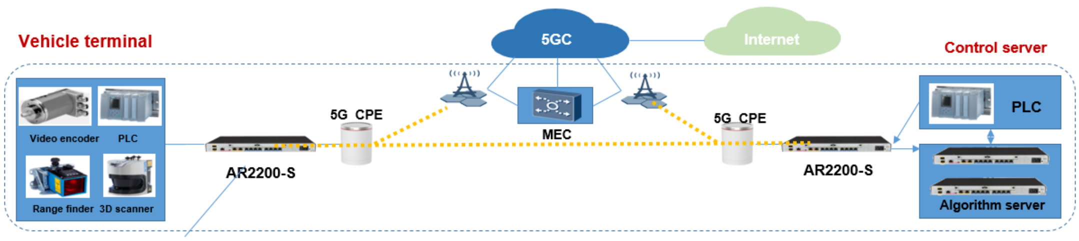


- The interaction latency increases significantly from 2 ms to 5 ms.

5.2.2 PLC-S7Comm

The following figure shows PLC service behavior of remote video control in the industrial park at site X.

Figure 5-7 PLC networking for remote video control at site X



S7Comm is a proprietary protocol of Siemens. This protocol is encapsulated in the TPKT and ISO-COTP protocols to enable the protocol data units (PDUs) to be transmitted through the TCP. It is used for PLC programming, data exchange between PLCs, access to PLC data from SCADA systems, and diagnostics.

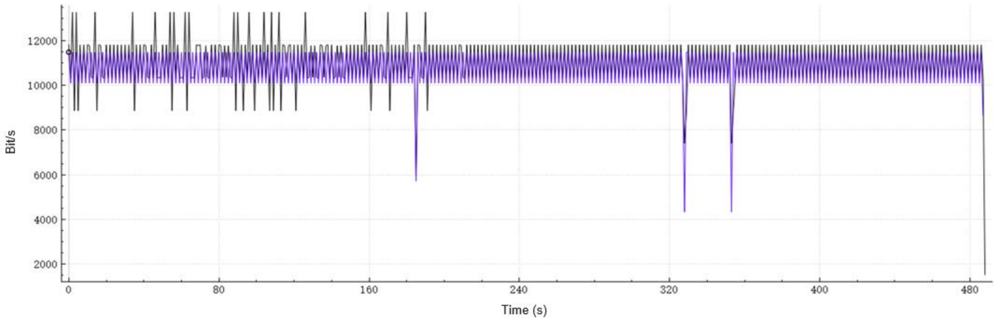
Analysis of captured PLC packets:

The Siemens S7Comm protocol is used.

No.	Time	Source	Destination	Protocol	ROSCTR	Function	The RTT	Length	Info
1	0.000000	192.168.2.171	192.168.2.172	GTP <TCP>				158	49175 → 102 [ACK] Seq=1 Ack=1 Win=4096 Len=0
3	0.124565	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var		183	ROSCTR:[Job] Function:[Read Var]
4	0.139485	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	0.014…	183	ROSCTR:[Ack_Data] Function:[Read Var]
5	0.185111	192.168.2.171	192.168.2.172	GTP <TCP>			0.045…	158	49175 → 102 [ACK] Seq=32 Ack=32 Win=4096 Len=0
7	0.224752	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var		188	ROSCTR:[Job] Function:[Write Var]
8	0.239524	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	0.014…	174	ROSCTR:[Ack_Data] Function:[Write Var]
9	0.294481	192.168.2.171	192.168.2.172	GTP <TCP>			0.054…	158	49175 → 102 [ACK] Seq=68 Ack=54 Win=4096 Len=0
11	0.324987	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var		183	ROSCTR:[Job] Function:[Read Var]
12	0.339458	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	0.014…	183	ROSCTR:[Ack_Data] Function:[Read Var]
13	0.390030	192.168.2.171	192.168.2.172	GTP <TCP>			0.050…	158	49175 → 102 [ACK] Seq=99 Ack=85 Win=4096 Len=0
15	0.524502	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var		183	ROSCTR:[Job] Function:[Read Var]
16	0.539456	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	0.014…	183	ROSCTR:[Ack_Data] Function:[Read Var]
17	0.590105	192.168.2.171	192.168.2.172	GTP <TCP>			0.050…	158	49175 → 102 [ACK] Seq=130 Ack=116 Win=4096 Len=0
19	0.625062	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var		188	ROSCTR:[Job] Function:[Write Var]
20	0.644432	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	0.019…	174	ROSCTR:[Ack_Data] Function:[Write Var]
21	0.694476	192.168.2.171	192.168.2.172	GTP <TCP>			0.050…	158	49175 → 102 [ACK] Seq=166 Ack=138 Win=4096 Len=0
23	0.734467	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var		183	ROSCTR:[Job] Function:[Read Var]
24	0.749436	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	0.014…	183	ROSCTR:[Ack_Data] Function:[Read Var]
25	0.785008	192.168.2.171	192.168.2.172	GTP <TCP>			0.035…	158	49175 → 102 [ACK] Seq=197 Ack=169 Win=4096 Len=0
27	0.924492	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var		183	ROSCTR:[Job] Function:[Read Var]
28	0.939445	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	0.014…	183	ROSCTR:[Ack_Data] Function:[Read Var]

> PPP Bridging Control Protocol Bridged PDU
> Ethernet II, Src: Siemens_56:b5:31 (28:63:36:56:b5:31), Dst: SiemensI_84:ac:d1 (e0:dc:a0:84:ac:d1)
> Internet Protocol Version 4, Src: 192.168.2.171, Dst: 192.168.2.172
> Transmission Control Protocol, Src Port: 49175, Dst Port: 102, Seq: 1, Ack: 1, Len: 31
> TPKT, Version: 3, Length: 31
> ISO 8073/X.224 COTP Connection-Oriented Transport Protocol
> S7 Communication

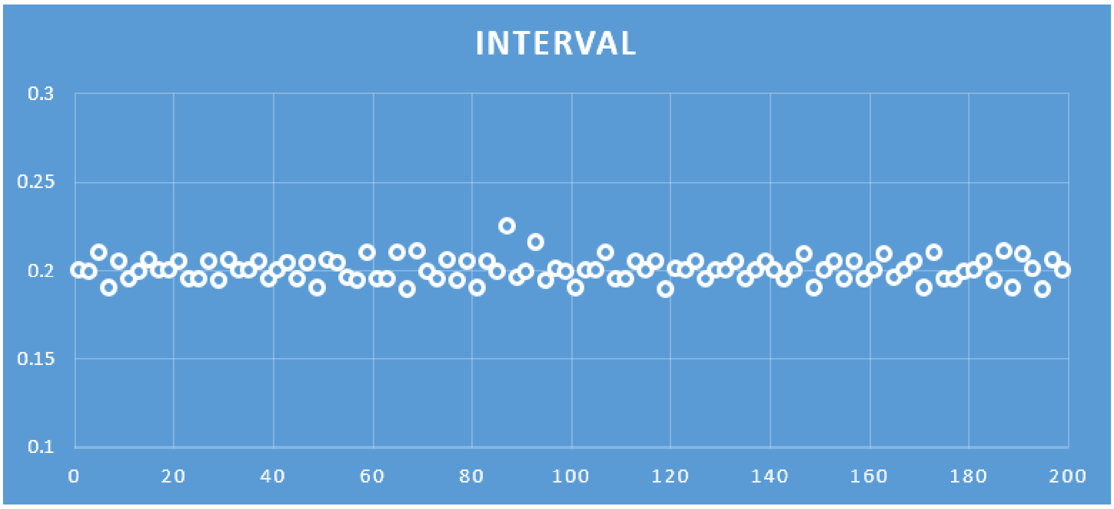
The following figure shows the data I/O diagram.



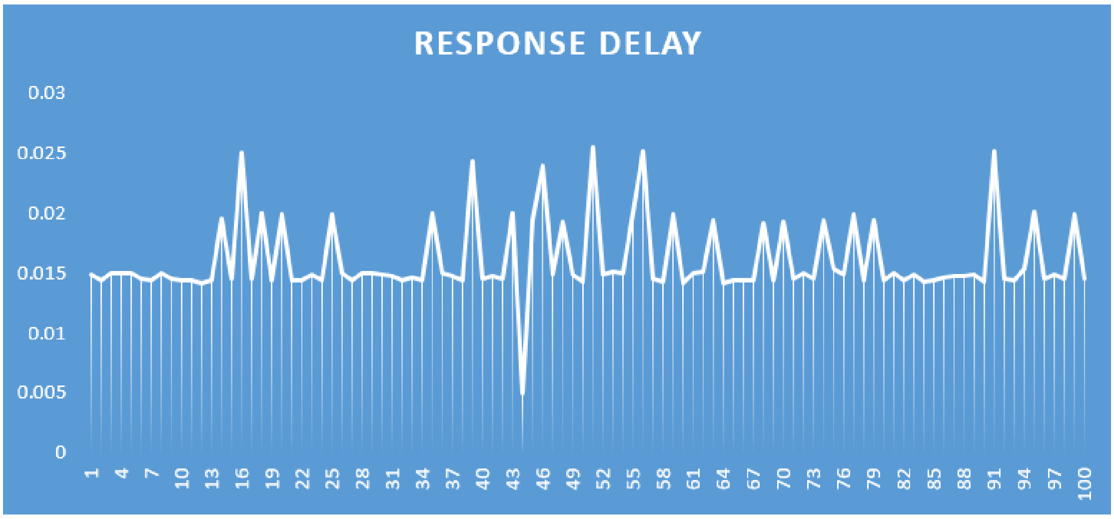
The Read Var function packets are as follows:

No.	Time	Source	Destination	Protocol	ROSCTR	Function	Lengt	Info
3	0.124565	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
5	0.139485	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
15	0.324987	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
17	0.339458	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
21	0.524502	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
23	0.539456	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
33	0.734467	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
35	0.749436	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
39	0.924492	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
41	0.939445	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
51	1.1299	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
53	1.14445	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
57	1.325067	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
59	1.339494	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
69	1.524465	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
71	1.539424	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
75	1.729958	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
77	1.744479	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
87	1.930045	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
89	1.944422	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
93	2.130011	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
95	2.144431	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
105	2.335203	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
107	2.34943	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
111	2.529973	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
113	2.544406	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]
123	2.724898	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Read Var	183	ROSCTR:[Job] Function:[Read Var]
125	2.744419	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Read Var	183	ROSCTR:[Ack_Data] Function:[Read Var]

The interval between request-and-response packet groups is 200 ms.



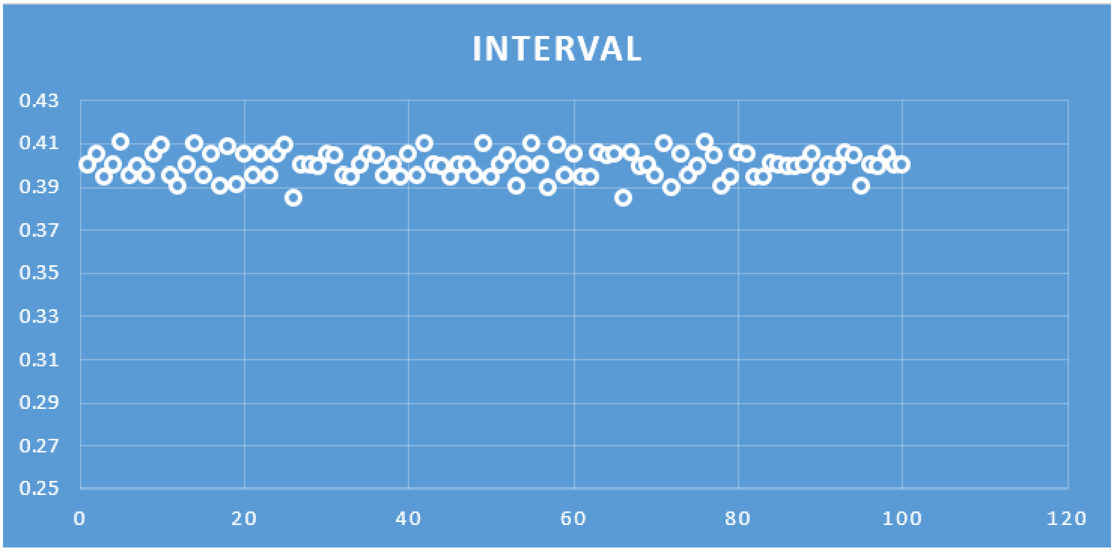
The average latency from request to response is about 15 ms, and fluctuates according to the network quality.



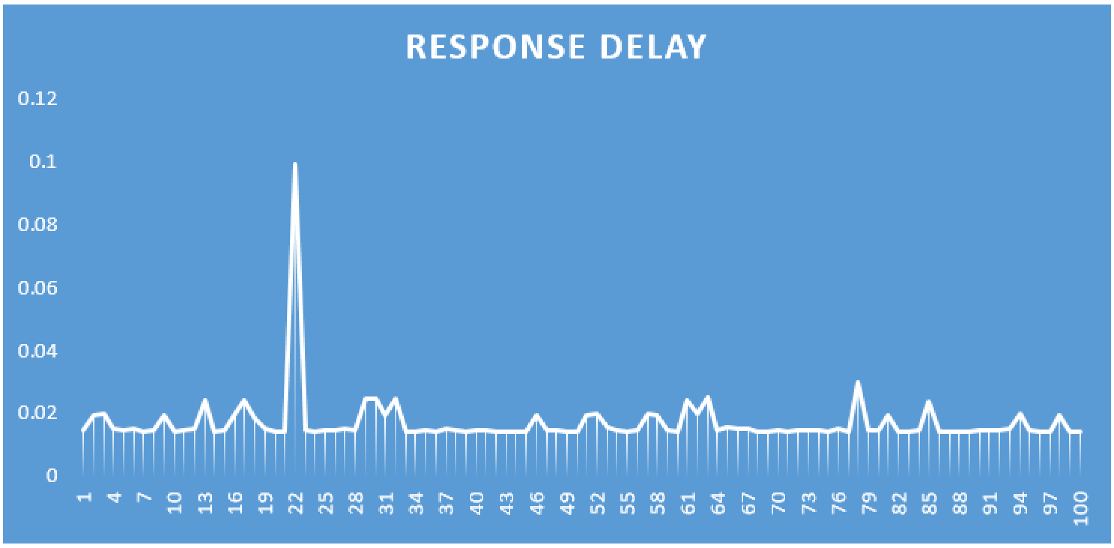
The Wright Var function packets are as follows:

No.	Time	Source	Destination	Protocol	ROSCTR	Function	Length	Info
9	0.224752	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
11	0.239524	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
27	0.625062	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
29	0.644432	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
45	1.029947	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
47	1.049974	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
63	1.424476	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
65	1.439474	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
81	1.824552	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
83	1.839426	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
99	2.234961	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
101	2.249934	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
117	2.630164	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
119	2.644408	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
135	3.029892	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
137	3.0444	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
153	3.42494	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
155	3.444448	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
171	3.830155	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
173	3.844442	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
189	4.239556	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
191	4.254412	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
207	4.634541	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
209	4.649694	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
225	5.024919	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
227	5.049388	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
243	5.425033	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]
245	5.439461	192.168.2.172	192.168.2.171	GTP <S7COMM>	Ack_Data	Write Var	174	ROSCTR:[Ack_Data] Function:[Write Var]
261	5.834822	192.168.2.171	192.168.2.172	GTP <S7COMM>	Job	Write Var	188	ROSCTR:[Job] Function:[Write Var]

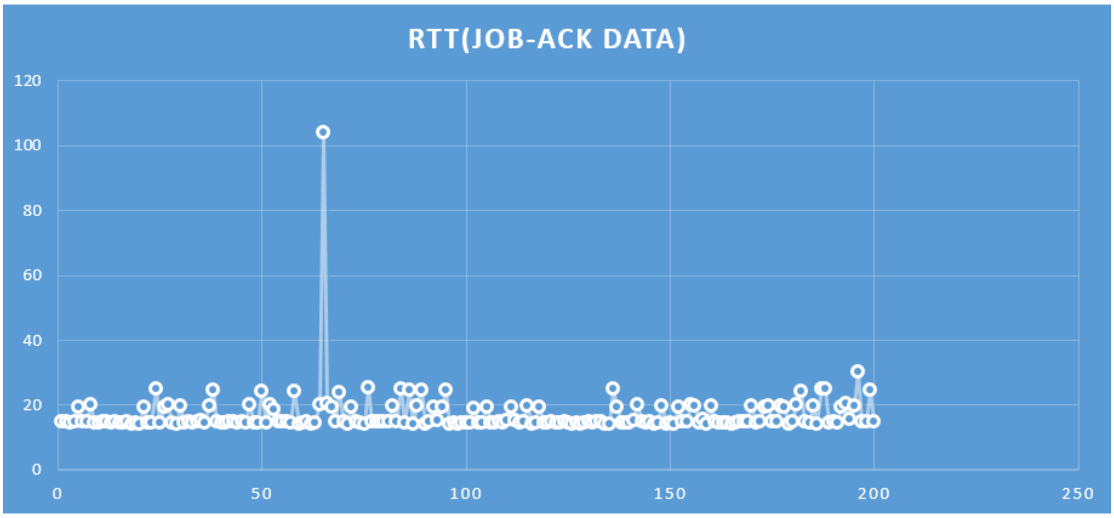
The interval between request-and-response packet groups is 400 ms.



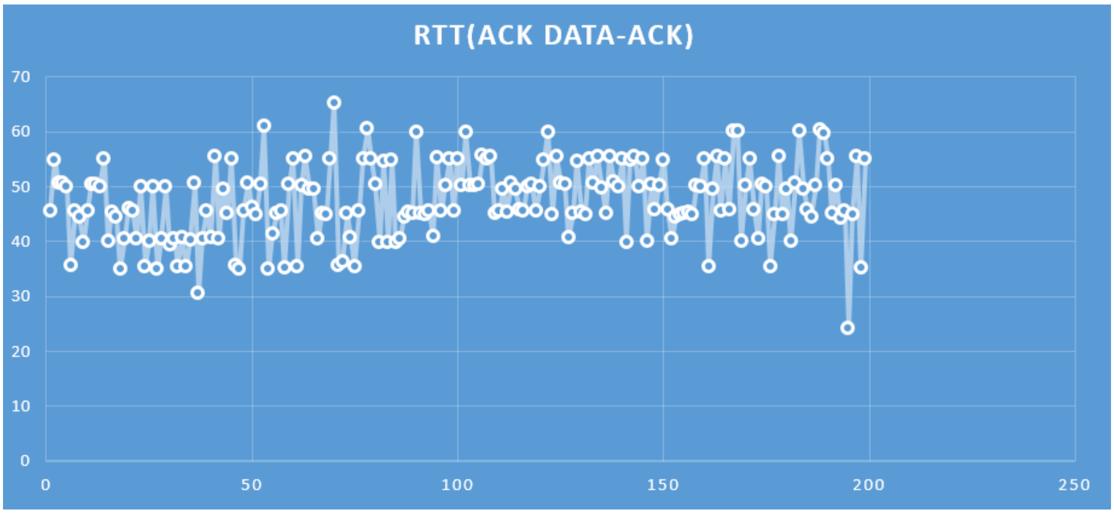
The average latency from request to response is about 17 ms, and fluctuates according to the network quality.



The RTTs in both directions can be measured because TCP is used. The RTTs are as follows:

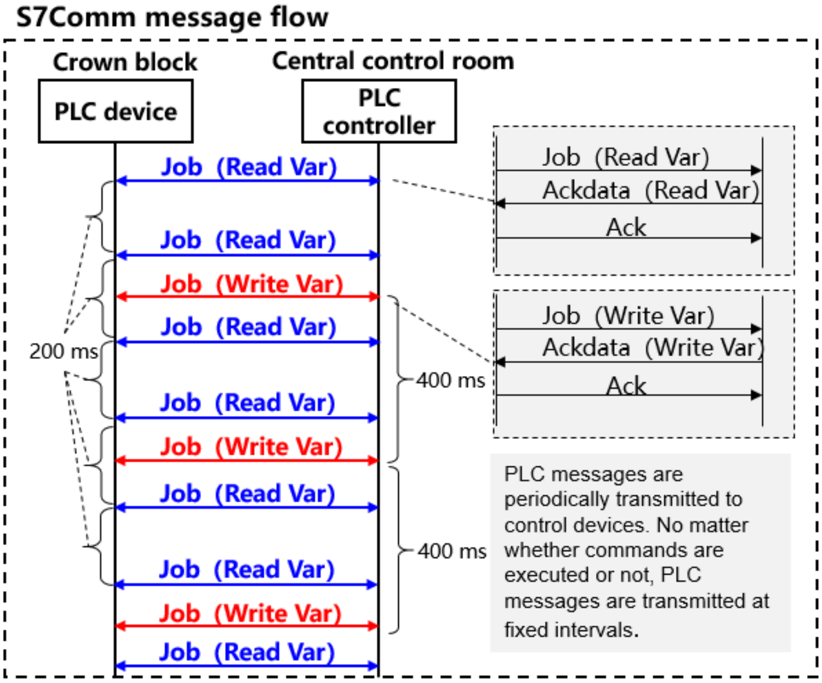


The average RTT is 17 ms. The RTT at one point is large, reaching 104 ms. This value corresponds to the high point of Response Delay in the preceding figure.



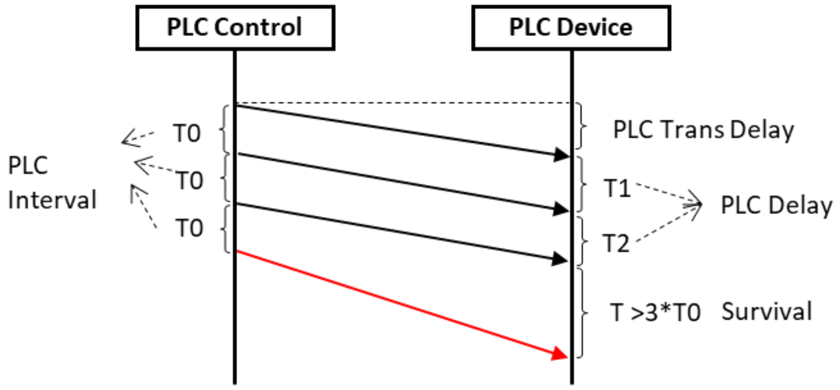
The average latency is 47 ms. Fluctuation is relatively small, ranging from 30 ms to 65 ms; this RTT reflects the network quality from ACK_DATA to ACK and cannot be covered by Response Delay.

According to the preceding analysis, the interactive behavior of the S7Comm in this scenario is shown in the following figure.



According to the S7Comm protocol, there is no specific protection mechanism for transmission latency. Therefore, even if the network transmission time is long, no alarm is generated (provided that transmission is stable). However, the protocol has a periodic status update mechanism for messages; each time a message is received, the periodic timer is updated.

As shown in the following figure, the message indicated by the black line reaches the peer end after a period of time, and if reached stably, the peer end does not trigger an alarm. However, if the network jitter is large and the message is not received for a long time, as shown by the message indicated by the red line, an alarm will be triggered.

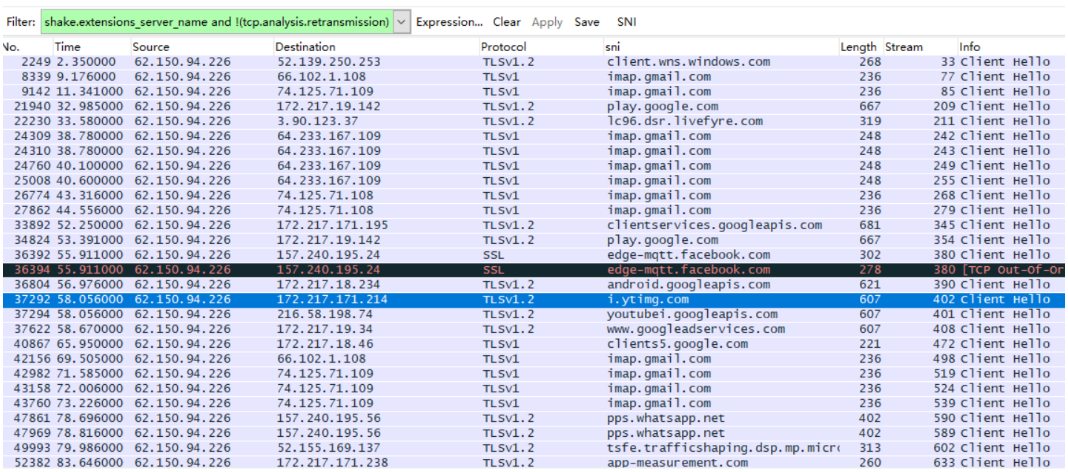


Based on the preceding analysis, the following indicators are supported:

Table 5-1 PLC message indicators

Indicator Name	Measurement Principles
Response Delay	Latency from Job to Ack Data
RTT (Job-Ack)	RTT from Job to Ack
RTT (Ack Data-Ack)	RTT from Ack Data to Ack
RTT Jitter	RTT fluctuation
Interval Delay	Period latency between jobs
Survival Times	The number of times that the period latency exceeds the emergency latency.

5.3 FWA Service



Filter: shake.extensions_server_name and !(tcp.analysis.retransmission) Expression... Clear Apply Save SNI

No.	Time	Source	Destination	Protocol	sni	Length	Stream	Info
2249	2.350000	62.150.94.226	52.139.250.253	TLSv1.2	client.wns.windows.com	268	33	Client Hello
8339	9.176000	62.150.94.226	66.102.1.108	TLSv1	imap.gmail.com	236	77	Client Hello
9142	11.341000	62.150.94.226	74.125.71.109	TLSv1	imap.gmail.com	236	85	Client Hello
21940	32.985000	62.150.94.226	172.217.19.142	TLSv1.2	play.google.com	667	209	Client Hello
22230	33.580000	62.150.94.226	3.90.123.37	TLSv1.2	lc96.dsr.livefyre.com	319	211	Client Hello
24309	38.780000	62.150.94.226	64.233.167.109	TLSv1	imap.gmail.com	248	242	Client Hello
24310	38.780000	62.150.94.226	64.233.167.109	TLSv1	imap.gmail.com	248	243	Client Hello
24760	40.100000	62.150.94.226	64.233.167.109	TLSv1	imap.gmail.com	248	249	Client Hello
25008	40.600000	62.150.94.226	64.233.167.109	TLSv1	imap.gmail.com	248	255	Client Hello
26774	43.316000	62.150.94.226	74.125.71.108	TLSv1	imap.gmail.com	236	268	Client Hello
27862	44.556000	62.150.94.226	74.125.71.108	TLSv1	imap.gmail.com	236	279	Client Hello
33892	52.250000	62.150.94.226	172.217.171.195	TLSv1.2	clientservices.googleapis.com	681	345	Client Hello
34824	53.391000	62.150.94.226	172.217.19.142	TLSv1.2	play.google.com	667	354	Client Hello
36392	55.911000	62.150.94.226	157.240.195.24	SSL	edge-mqtt.facebook.com	302	380	Client Hello
36394	55.911000	62.150.94.226	157.240.195.24	SSL	edge-mqtt.facebook.com	278	380	[TCP Out-Of-Or
36804	56.976000	62.150.94.226	172.217.18.234	TLSv1.2	android.googleapis.com	621	390	Client Hello
37292	58.056000	62.150.94.226	172.217.171.214	TLSv1.2	i.ytimg.com	607	402	Client Hello
37294	58.056000	62.150.94.226	216.58.198.74	TLSv1.2	youtubei.googleapis.com	607	401	Client Hello
37622	58.670000	62.150.94.226	172.217.19.34	TLSv1.2	www.googleadservices.com	607	408	Client Hello
40867	65.950000	62.150.94.226	172.217.18.46	TLSv1	clients5.google.com	221	472	Client Hello
42156	69.505000	62.150.94.226	66.102.1.108	TLSv1	imap.gmail.com	236	498	Client Hello
42982	71.585000	62.150.94.226	74.125.71.109	TLSv1	imap.gmail.com	236	519	Client Hello
43158	72.006000	62.150.94.226	74.125.71.109	TLSv1	imap.gmail.com	236	524	Client Hello
43760	73.226000	62.150.94.226	74.125.71.109	TLSv1	imap.gmail.com	236	539	Client Hello
47861	78.696000	62.150.94.226	157.240.195.56	TLSv1.2	pps.whatsapp.net	402	590	Client Hello
47969	78.816000	62.150.94.226	157.240.195.56	TLSv1.2	pps.whatsapp.net	402	589	Client Hello
49993	79.986000	62.150.94.226	52.155.169.137	TLSv1.2	tsfe.trafficshaping.dsp.mp.micr	313	602	Client Hello
52382	83.646000	62.150.94.226	172.217.171.238	TLSv1.2	app-measurement.com	260	633	Client Hello

Ethernet: 1 | Fibre Channel | FDDI | IPv4: 271 | IPv6 | IPX | JXTA | NCP | RSVP | SCTP | TCP: 1194 | Token Ring | UDP: 302 | USB | WLAN

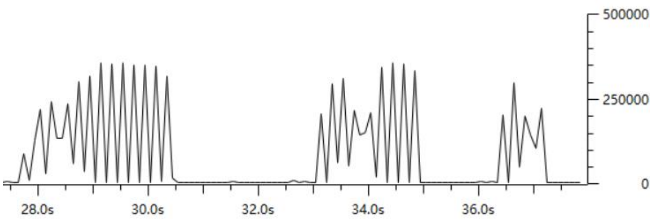
TCP Conversations

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration
62.150.94.226	49770	74.125.71.109	imaps	3 488	2 987 260	1 190	87 884	2 298	2 899 376	71.886000000	24.2040
62.150.94.226	50415	64.233.166.109	imaps	3 126	2 602 814	1 106	80 454	2 020	2 522 360	86.441000000	17.2100
62.150.94.226	49776	74.125.71.109	imaps	2 846	2 515 404	964	70 478	1 882	2 444 926	146.256000000	15.3550
62.150.94.226	51114	74.125.71.108	imaps	2 942	2 272 560	972	75 094	1 970	2 197 466	44.436000000	61.4630
62.150.94.226	51117	74.125.71.108	imaps	2 900	2 268 970	948	73 330	1 952	2 195 640	106.856000000	19.8490
62.150.94.226	55893	64.233.167.109	imaps	2 868	2 155 552	992	90 218	1 876	2 065 334	40.480000000	62.2650
62.150.94.226	49764	74.125.71.109	imaps	2 202	1 964 966	744	53 526	1 458	1 911 440	0.039000000	11.3020
62.150.94.226	49766	74.125.71.109	imaps	1 632	1 074 532	566	45 552	1 066	1 028 980	11.211000000	135.1260
62.150.94.226	59981	157.240.195.56	https	1 050	1 018 220	348	31 464	702	986 756	78.586000000	65.6560
62.150.94.226	49608	172.217.18.36	https	1 310	1 006 850	466	51 088	844	955 762	136.736000000	16.6700
62.150.94.226	53274	77.238.180.12	https	1 556	896 698	642	126 032	914	770 666	17.466000000	144.1300
192.168.4.254	65291	192.168.1.59	citriximac	7 172	1 024 768	4 006	301 368	3 166	723 400	0.606000000	160.8830
62.150.94.226	59984	64.233.166.108	imaps	772	668 860	258	19 886	514	648 974	137.851000000	3.7200
62.150.94.226	53731	143.204.106.100	https	756	663 038	298	41 108	458	621 930	147.011000000	11.5400
62.150.94.226	50414	64.233.166.109	imaps	1 284	562 610	452	39 158	832	523 452	85.931000000	53.3470
62.150.94.226	61323	66.102.1.108	imaps	1 048	469 826	360	31 422	688	438 404	69.390000000	86.9210
62.150.94.226	61319	66.102.1.108	imaps	1 040	467 546	364	31 488	676	436 058	9.055000000	60.4190
62.150.94.226	55892	64.233.167.109	imaps	1 138	461 346	404	40 436	734	420 910	39.970000000	63.0160

According to code stream analysis, the FWA service feature is the same as the B2C service flow feature. Therefore, the B2C service can be used for evaluation. The difference is that the source IP address is the IP address of the CPE and cannot be distinguished by user level. However, different service CPEs are allocated with different port numbers, and as such, the flow data of each service can be viewed separately from the 5-tuple layer.

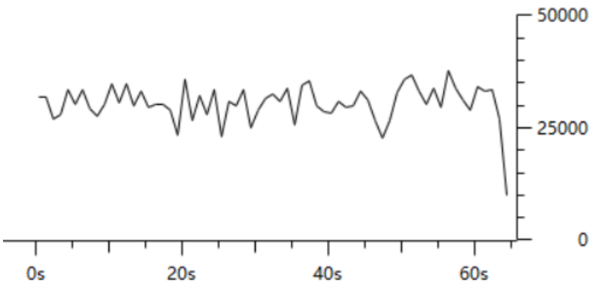
[Enterprise Internet service]

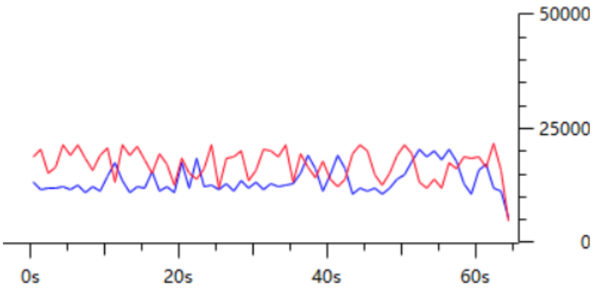
- Gmail receiving behavior



Download code streams. The downlink rate is about 2.5 Mbps, and comprises most of the traffic of enterprise Internet.

- VoIP service code stream



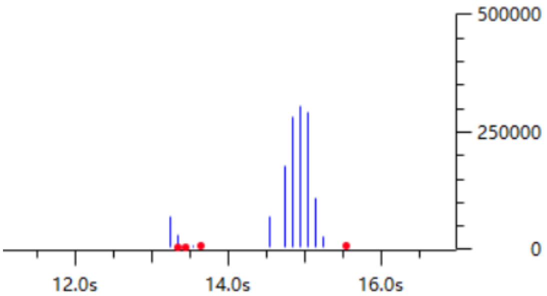


The uplink rate is shown in red, and the downlink rate in blue. The average uplink rate is 18 kbps, and the average downlink rate is about 15 kbps.

According to packet length statistics, packet sizes range from 80 bytes to 159 bytes, with more than 90% ranging from 50 bytes to 200 bytes. The packets are small; the average packet size is 120 bytes, and the average packet interval is 32 ms. Both are VoIP service features.

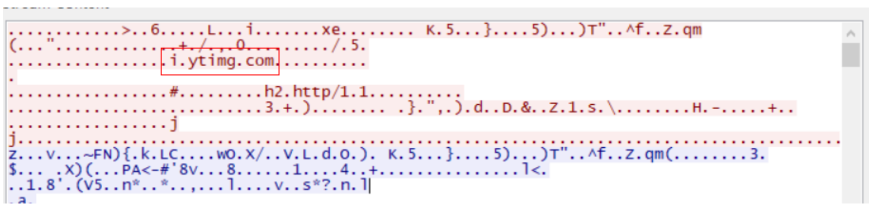
Topic / Item	Count	Rate (ms)	Percent
⊟ Packet Lengths	2053	0.031921	
0-19	0	0.000000	0.00%
20-39	0	0.000000	0.00%
40-79	413	0.006422	20.12%
80-159	1052	0.016357	51.24%
160-319	588	0.009143	28.64%

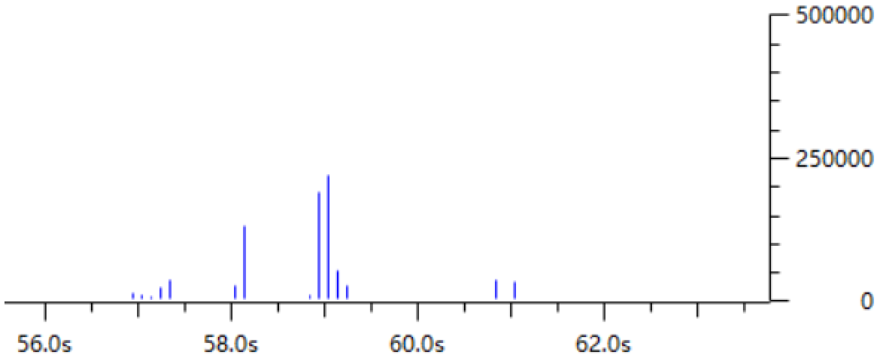
- WhatsApp download data code stream



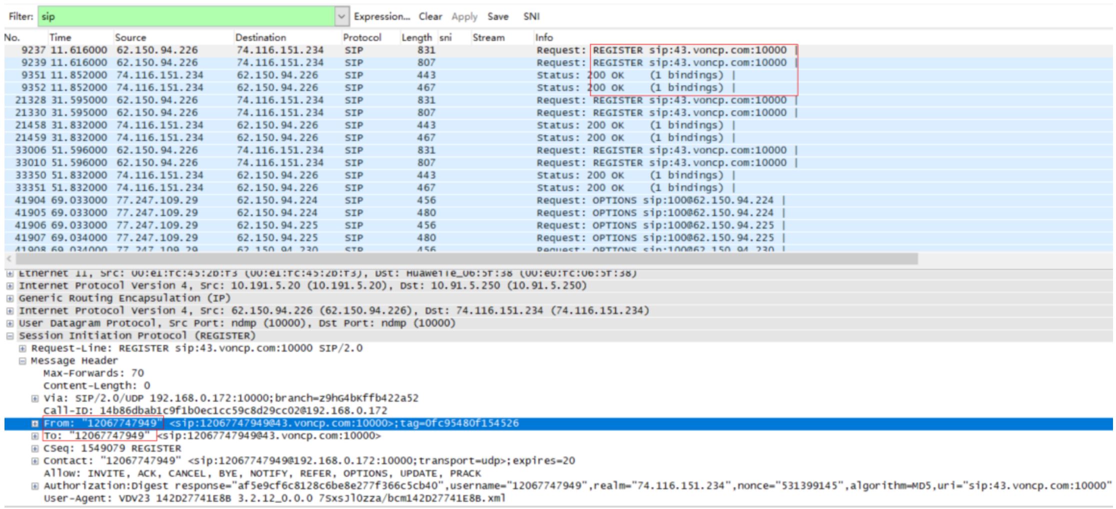
The uplink code stream is shown in red, and downlink code stream in blue. The downlink rate is 2.5 Mbps.

- Web browsing data flow





The code stream shows the SIP registration process, which is probably an IP phone of the IP PBX registering with the server. The SIP message is not encrypted, and the source and destination numbers are visible.



6 5G ToB Service Modeling Framework

6.1 ToB & B2C Modeling Differences

Table 6-1 B2C and ToB modeling analysis dimensions

Analysis Dimension/Service Type	B2C Experience Modeling	ToB Experience Modeling
Quality commitment	Operators do not make SLA commitments to individual users.	Operators make SLA commitments to enterprise customers.
Satisfaction	Complaints and churn occur following poor single-user experience: Individual experience modeling is important.	No complaints from things but if the overall SLA does not meet the requirements, customers may claim for compensation based on contracts. Group experience modeling is more important.
Troubleshooting	B2C users can rectify the fault by themselves by powering off, calling the assistance hotline, or consulting associates.	ToB users cannot. Once a fault occurs, the system will be suspended.
Traffic model	B2C services are bursts (long-time or short-time data transmission).	ToB services are continuous or involve regular bursts. Services are always online.
Software application	B2C software systems are dominated by few providers.	The ToB-oriented software systems present long tail characteristics.
Network characteristics	Shared networks cannot guarantee differentiated QoS for different services.	SA networking and slicing: ensures differentiated QoS requirements of different services.
Technical challenges	Encrypted identification, experience modeling, and traffic explosion	E2E QoS measurement (UDP) and dynamic QoS guarantee
KPI difference	Throughput and latency	In addition to throughput and latency, consider energy consumption, network resource usage, and abnormal distribution of objects.
Optimization points	Wireless RF quality, CN-SP route/rate limiting/packet loss	Wireless RF quality, network structure adjustment, and resource allocation policy

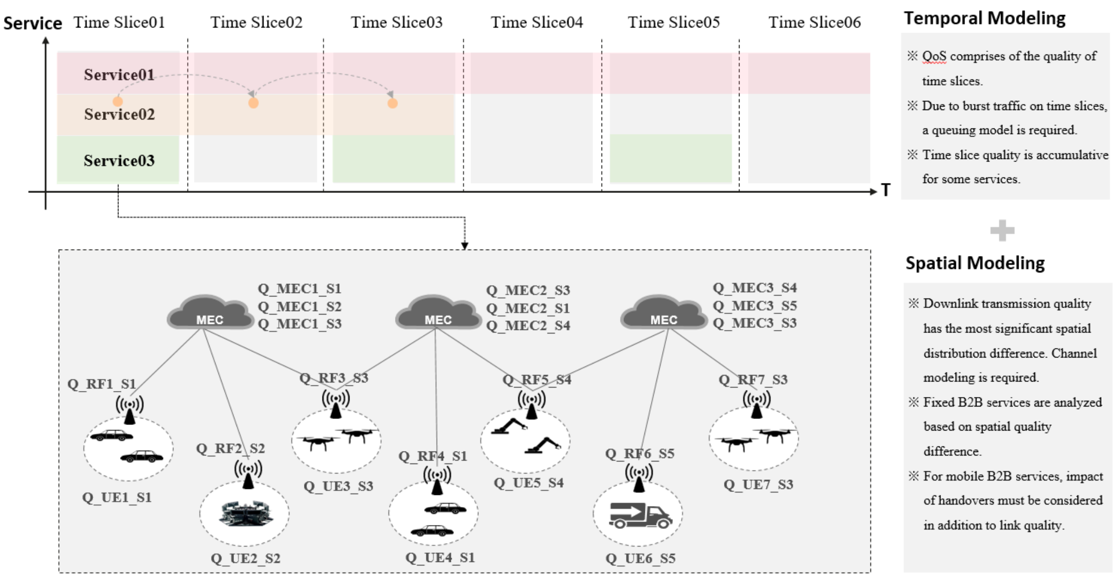
Table 6-2 B2C and ToB modeling differences

B2C Experience Modeling	ToB Experience Modeling
PSPU individual experience modeling	Group quality modeling for "things"
Ensuring the monopolistic applications, categorized experience modeling	Too many application scenarios, customized modeling + general categorized modeling
Strive for ultimate user experience	Optimal balance between network resources and experience
Precise QoS measurement (for example, RTT 99.9% precision)	Precise measurement + AI-based quality prediction (with confidence)
Experience evaluation and demarcation are the driving force of service solutions.	Experience assurance is the driving force of service solutions.

6.2 ToB Modeling Method Exploration

6.2.1 Fine-grained Spatio-temporal Modeling

Figure 6-1 ToB QoS and space-time relationship



ToB QoS modeling must be based on time and space.

- In the time domain, event-driven instantaneous traffic impact or packet quantity impact causes burst latency and packet loss. As radio devices and core network/bearer network devices perform instantaneous queuing, the queuing mechanism of different services needs to be modeled to quantify the QoS impact such as the latency, packet loss, and bandwidth of service processing within a time slice.

- In the space domain, rapid changes in channel quality caused by mobility and communication location changes are quantified in channel quality modeling. MEC resource allocation also impacts the overall ToB service quality. If MEC planning and resource allocation are properly performed, the MEC processing latency and application layer processing latency are not closely related to the location. However, this parameter is not considered.
- Spatio-temporal modeling: First, the quality of a single time slice is modeled, with the impact of the queuing model and the channel model on the quality fully considered in the single time slice. For services that span multiple time slices, check whether the quality of the time slices is related. If they are related, use the state change method (such as Markov) for associated evaluation. If the quality of time slices is independent, the weighted average or the moving weighted average considering the near-end effect can be used to evaluate the comprehensive service quality.

The average performance alone is insufficient for ToB service quality evaluation. Transient poor quality (burst latency, burst congestion, burst packet loss, and burst jitter) is generated at the time slice level. Different objects in the same time slice represent different positions in space and may change over time. Assume that a location of an object in a same time slice is fixed. It should be noted that service quality varies greatly at different locations, that is, the final comprehensive quality is a two-dimensional function of time and space.

The following uses end-to-end latency as an example. Assume that the latency (D) is a two-dimensional function of time and space:

$$D(x_i, y_j), i \in [1, N], j \in [1, M]$$

In the formula, x_i is the current time slice, the maximum quantity of time slices in the evaluation period is N, y_j is the current object, and each object has a different location, thus represents a different spatial location, and the maximum quantity of objects is M.

$$f(x, y) = \iint_{i=1,j=1}^{i=N,j=M} D(x_i, y_j)dxdy$$

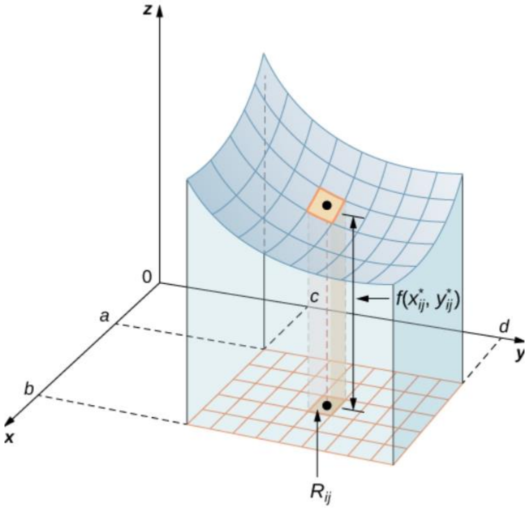
$f(x, y)$ is the double integral of D in the space-time dimension and represents the accumulated measurement of the latency, which is presented as the volume of the blue area in the figure.

$$f^*(x, y) = \iint D^*(x_i, y_j)dxdy\, , D^*(x_i, y_j) > D_{boundary}$$

$f^*(x, y)$ indicates the sample integral of the latency that exceeds the boundary.

$$r(x, y) = P(D > D_{boundary}) = \frac{f^*(x, y)}{f(x, y)} \times 100\%$$

Figure 6-2 Double integral representation of spatial-temporal distribution of mass



In a measurement period, the E2E latency of all objects can be measured using $f(x, y)$ and $r(x, y)$.

6.2.2 Scenario-based Event-driven Modeling

In ToB networks, wireless sensors are used frequently, and all applications that use in-situ sensors strongly depend on their proper operation, which is difficult to ensure. These sensors are usually cheap and prone to failure. For many tasks, sensors are used in harsh weather conditions, making them more vulnerable to damage. In addition, industrial devices have high requirements on reliability. Common faults can be detected by alarms in the tenant system. However, hidden faults are difficult to detect due to external factors or aging. If they are not handled in a timely manner, faults gradually occur, reducing the SLA. Therefore, the pre-detection and pre-analysis of the abnormal behavior of objects are significant to the preventive management of enterprises. And because of the number and variety of things, the detection process must be automated, scalable, and fast enough for real-time streaming data.

In conclusion, machine learning and heuristic learning-based anomaly detection technologies will play an increasingly important role in various future 5G IoT applications.

Anomaly Detection in Intelligent Inhabitant Environment

In intelligent inhabitant environment, embedded sensor technology plays a major role in monitoring occupants' behavior. The inhabitants interact with household objects, and embedded sensors generate time-series data to recognize performed activities. Generated sensor data is very sparse, because the sensor values change when the inhabitant interacts with objects. The need for robust anomaly detection models is essential in any intelligent environment.

- Statistical methods in intelligent inhabitant environment

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Dissimilarity Measures [17]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	State and motion sensors	Smart home	Distance index
Percentiles [18]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Motion and door sensors	Senior's inhabitant home	False alert rate and sensitivity
Gaussian Mixture Model [19]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Occupancy, motion and door sensors	UK sheltered housing scheme	Histogram Visualization
Hierarchical Markov Model [20]	Binary and continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Embedded state sensors in Home	MIT Placelab smart home	Accuracy
Switching Hidden Semi-Markov Model [21]	Continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Video cameras	Smart kitchen	Average Accuracy
Bayesian Model [22]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Binary sensors, motion sensors and pressure sensors	Smart home	Precision, Recall, and F-measures

- Machine learning methods in intelligent inhabitant environment

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Single Class Support Vector Machine (SVM) [16]	Binary data	Collective anomaly	Supervised learning over the patterns of normal behavior	State sensors deployed in living room, kitchen and dining area	Smart home	Type I and II error, Precision, Recall, F - Measure
Multi-class SVM [24]	Continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	Accelerometer and gyroscope sensors	Wearable gadgets	Accuracy, Precision, Sensitivity
Support Vectors [25]	Continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Infrared motion sensors	Ubiquitous Health-care House	Positive predictive value
Principal Component Analysis (PCA) and Fuzzy Rule-based System [26]	Binary and continuous data	Collective anomaly	Supervised learning over the patterns of normal behavior	state and motion sensors	Smart home	Hotelling's T2 and Square prediction error (SPE)
Kernel Nonlinear Regression and SVM [27]	Continuous data	Contextual anomaly	Supervised learning over the patterns of normal behavior	Light, temperature, microphone, accelerometer, magnetometer	Human activities	False positive rate
Convolutional neural network (CNN) and Recurrent Neural Network (RNN) [28]	Continuous data and Images	Collective anomaly	Supervised learning over the patterns of normal behavior	Microwave sensor and video camera	Smart home	Accuracy and Mean Absolute Error (MAE)

Anomalous Behavior in Intelligent Transportation System

- Statistical methods in intelligent transportation systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Kernel Feature Space and PCA [30]	Multi-dimensional time-series	Collective anomaly	Supervised learning over Telemetry data from space station	System components (Soft sensors) in spacecraft	Aerospace	False alarms
K-means and GMM [31]	Numerical, categorical, and binary.	Contextual anomaly	Unsupervised learning over EuroFOT database	403 parameters of hardware and software sensors,	Anomaly detection for road traffic	Colorscales (red for anomalies)
Structured sparse subspace learning [32]	Continuous data.	Contextual anomaly	Supervised learning on Thor Flight 107 and 111 flight data	flight-critical sensors	Anomaly detection for flight safety	Accuracy and ROC

- Machine learning methods in intelligent transportation systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Decision tree (C4.5) and Fusion Model [33]	Unstructured data in the form of reports	Collective anomaly	Supervised learning over historical data collected over 5 years	Operational data of engines of A-320	Aerospace	Accuracy and Error rate
Multiple Kernel Learning [34]	Discrete and continuous data	Collective anomaly	Supervised learning over the pattern of flights	Different 160 parameters of FOQA dataset	Aviation safety	Accuracy
Extreme Learning Machines [35]	Continuous data	Collective anomaly	Supervised learning on 43000 flights data	Radar, aircraft trajectories and nearby aircraft distance	Aviation safety	Area under the curve
Reinforcement Learning [36]	Continuous data	Point anomaly	Supervised learning on drone captured data	Temperature sensor	Unmanned aerial vehicles	Accuracy
Regression Model [37]	Continuous data	Contextual anomaly	Supervised learning on historical data	Software Logs	Air Traffic Control systems	Precision, Recall and Accuracy
Support Vector Machine [38]	Continuous data	Contextual anomaly	Supervised learning on historical data	GPS and Accelerometer	Air Driving Patterns and Road Anomalies	Accuracy
Deep Convolutional Neural Network [39]	Image data	Point anomaly	Supervised learning on CTIV platform used to collect the images	Camera	Railway track	Accuracy
Autoencoders [40]	Continuous data	Point anomaly	Unsupervised learning on data from B777-200 of civil plane	Gas turbine engine	Aerospace	Precision, Recall, F1 Score

Anomalous Behavior in Smart Objects

The smart object is a fast-growing area to connect multiple objects together and enable communication between them. It collects valuable data that can be a source of information and knowledge for a wide range of applications. During our research, we found the following statistical and machine learning literature that is aligned with our research questions and search criteria

- Statistical methods in smart objects

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Threshold Level [41]	Continuous data	Point anomaly	Supervised learning on patterns of the food in trash bin	Gas sensors	Trash bin	No performance measures
Analysis of variance (ANOVA) [42]	Continuous data	Collective anomaly	Supervised learning on the driving section	Vehicle data from engine, fuel, gear and steering wheel.	Abnormal behavior of the vehicle	Normal vs abnormal data
Latent Correlation Probabilistic Model [43]	Continuous data	Collective anomaly	Supervised learning on patterns of concrete trucks	Engine speed and pump speed	Concrete pump trucks	Precision, Recall, area under the ROC
Expectation Maximization [44]	Binary data	Point anomaly	Unsupervised learning on switches patterns	Light switches	Smart nursing homes	False alarm rate

- Machine learning methods in smart objects

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Rule-based System [46]	Continuous and image data	Collective anomaly	Supervised learning	Temperature, humidity, light intensity, audio and image data	Efficient energy consumption in school building	Number of conflicts and Miss detection rate
Belief-rule-based Association Rules [45]	Continuous data	Collective anomaly	Supervised learning on patterns of incomplete and vague data	Temperature and rain gauge	Flood prediction	Area under the curve
Cluster heat maps [48]	Continuous data	Contextual anomaly	Supervised learning on patterns of smart cities IoT data	Measurement of total power, reactive power, phase angle, light, temperature, motion sensor, noise level, and vibration	Smart City	Normal vs abnormal comparisons
Temporal Clustering Technique [47]	Continuous data	Collective anomaly	Supervised learning on 8200 parking data	On-street parking spaces and gate controlled sensors	City parking of San Francisco	Correlations observations
Spatio-temporal Framework [49]	Continuous data	Collective anomaly	Supervised learning over 50 devices	Temperature, humidity sensor	Environmental data analysis of Taipei	Accuracy
Features Extraction and Visualization [50]	Continuous data	Collective anomaly	Supervised learning on patterns of smart cup usage	Raspberry Pi Zero, 9-DoF IMU, liquid level sensor, and force sensing resistors	Smart cup	No comparison

Anomalous Behavior in Healthcare Systems

Anomaly detection, analysis, and prediction are considered a revolution in redefining health care systems. In such systems, a clear impact can be seen on health management and wellness to improve quality of life and remote monitoring of chronic patients. Such systems pose a great challenge to reducing the generation of false alarms. In our systematic literature survey, we have found sufficient approaches and methods to identify anomalous behavior of sensors, humans, or machines in healthcare environment.

- Statistical methods in healthcare systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
RMS Graph [51]	Continuous data	Contextual anomaly	Supervised learning on wearable patterns	Accelerometer	Seizure complexities	Threshold configuration
Dynamic Time Warping (DTW) and Density Functions [53]	Continuous data	Contextual data	Supervised learning on Physionet PPG signals	Photoplethysmogram (PPG) signals	Cardiac behavior	Precision, Recall, Specificity
Autoregressive integrated moving average (ARIMA) [56]	Textual data	Point anomaly	Electronic health records	Electronic health records	Clinical decision support system	No performance measure
Hidden Markov Model [54]	Continuous data	Contextual anomaly	Supervised learning on eWALL Consortium parameters	Wearables Fitbit	Sleep analyzer	No performance measure
Hidden Markov Model [52]	Continuous data	Point anomaly	Supervised learning on glucose level	Medical devices insulin tolerance test	Blood glucose level behavior	Precision and Recall
Spectral Coherence Analysis [55]	Continuous data	Point anomaly		Accelerometer Sensor	Gait freezing in Parkinson's disease	Sensitivity

- Machine leaning methods in healthcare systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
HEAL Model [57]	Binary data	Contextual anomaly	Supervised learning on smart home data	Tap sensors on the objects	Abnormal behavioral	No performance measures
Incremental learning algorithm [58]	Continuous data	Contextual anomaly	Supervised learning on the dataset from UCI repository	Heart rate and blood pressure measuring sensor	Anomaly pattern in the health records	Actual vs predicted values
Transductive Transfer Learning [60]	Continuous data	Contextual anomaly	Semi-supervised learning on arrhythmias dataset	ECG monitoring sensor	Electrocardiogram abnormalities	G-mean
A Graph-based Approach [59]	Textual data	Contextual anomaly	Supervised learning on Hospital dataset	Access logs	Anomalies in electronic medical record	ROC curves
Support Vector Machine [61]	Textual anomaly	Contextual anomaly	Supervised learning on EHR	Data logs	Patient-management actions	True alert rate

Anomalous Behavior in Industrial Systems

In industrial systems, the design and development of anomaly detection methods are crucial to reduce the chance of unexpected system failures. It has been found that the developed methods for anomaly detection have been successfully applied to predictive and proactive maintenance. Such methods are widely used to improve productivity performance, save machine downtime, and analyze the root causes of faults.

- Statistical methods in industrial systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Temporal Logic [62]	Continuous data	Point anomaly	Supervised learning on 50 collections of traces	air compressor motor speed	Fuel cell vehicle	Misclassification rate
Correlation Analysis [63]	Continuous data	Contextual anomaly	Supervised learning on 5 machines data	Tap sensors on generator	Electric generators in factories	Correlation coefficients
Density Function Model [64]	Continuous data	Point anomaly	Supervised learning on 24 solar panels data	Electric current data	Solar power generation systems	ROC curves
Markov chain [65]	Continuous data	Point anomaly	Supervised learning on pressure sensor data	Pressure sensor	Oil pipeline	Accuracy

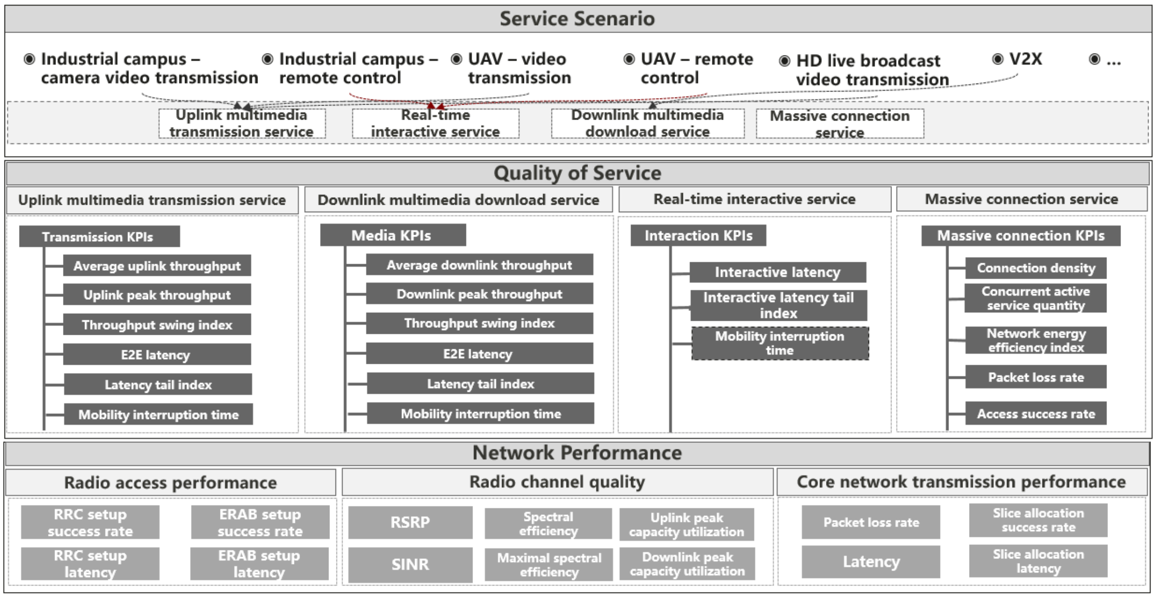
- Machine leaning methods in industrial systems

Methods	Nature of Data	Type of Anomaly	Availability of Data	Type of Sensors	Application Area	Evaluation Criteria
Extreme Learning Machines [69]	Continuous data	Point anomaly	Supervised learning on combustor chambers exhaust data	Temperature sensor	Power plant operations	ROC curves
Multivariate Clustering [71]	Continuous data	Contextual anomaly	Supervised learning on real-world sensory data	Sensor data from the electrical, water, and gas systems	Reduce electricity waste	Misclassification rate
Clustering [67]	Continuous data	Contextual anomaly	Unsupervised learning on five floor building	Temperature sensor	HVAC System in Smart Buildings	False alarms
One-Class SVM [66]	Continuous data	Point anomaly	Camshaft revolutions	A generic DAQ card (NI-6143)	An industrial seal machine	
Neural Network [68]	Continuous data	Contextual anomaly	Supervised learning on the data of thermal power plant	Superheated steam temperature, flow and steam cooling water flow.	Thermal power plant	Room mean square error
Conditional Gradient boosting decision tree (GBDT) [70]	Continuous data	Contextual anomaly	Supervised learning on the data of wind turbine	150 measurement parameters of wind turbine.	Wind turbine	Accuracy

6.3 ToB Modeling Frame

6.3.1 Indicator-driven Modeling Framework

Figure 6-3 ToB indicator-driven modeling framework analysis



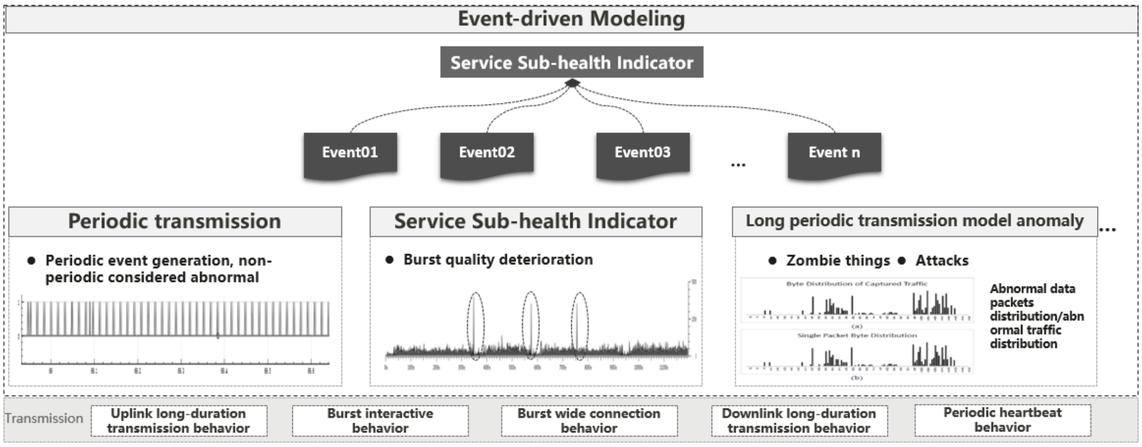
ToB service modeling framework:

- Service scenarios: The various ToB service scenarios can be generally categorized into three types: eMBB, uRLLC, and mMTC. Based on the understanding of service requirements of the current 5G project, from the perspective of network requirements, the services can be classified into uplink multimedia transmission services, downlink multimedia transmission services, real-time interactive services, and wide connection services. Currently, high-mobility services are not seen in projects. Theoretically, high-mobility services are a special scenario of multimedia services and real-time interactive services.
- Service quality: Different service types have different requirements on networks. Differentiated indicator systems are recommended.
- Network performance: Network is the foundation of service quality. Network performance can be classified into radio access performance, radio channel quality, and core network transmission performance.

Due to radio resource preemption and resource insufficiency, neither 4G nor 5G networks can meet each user's quality requirement. The SLA is not specific to individual users, rather, it is specific to the entire network.

6.3.2 Event-driven Modeling Framework

Figure 6-4 ToB event-driven modeling framework



This document describes the event-driven modeling method, which touches upon problems that cannot be covered by traditional latency/rate indicators. This method is from the perspective of identifying anomalies and analyzing problem types through big data and clustering analysis, rather than the perspective of PSPU experience modeling. To indicate the overall poor quality of a service behavior, you can set the service sub-health index as the comprehensive service quality evaluation indicator affected by anomalies.

General network data transmission behaviors can be classified into the following types:

- Uplink long-duration transmission behavior
- Downlink long-duration transmission behavior
- Burst wide connection behavior
- Burst small packet interaction
- Periodic heartbeat behavior

There is no universial anomaly detection solution that can help define the pattern of anomalies based on the service scenario and possible symptoms. Instead, case-by-case definition of anomaly events based on service characteristics is required. For development of platform products, consider the template, invoking mechanism, and upper-layer statistical indicator calculation model defined for abnormal events, which can be fixed to form product capabilities of edge and central nodes. However, behavior analysis and anomalies definition based on service behavior analysis by service delivery experts are required in frontline projects.

7 5G ToB Service Indicator System

7.1 Uplink Multimedia Transmission Service

7.1.1 Impact Factor

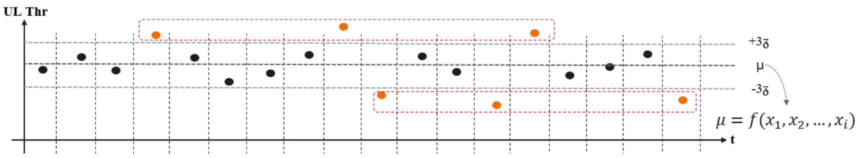
Multimedia transmission services include the following types:

1. For uplink real-time streaming media transmission services, such as live stream download and video surveillance, videos are transmitted in real time based on the video quality, such as the frame rate, bit rate, and resolution. The network transmission rate must meet the bit rate requirements, while the transmission latency must meet certain requirements.
2. Uplink multimedia message services, such as voice messages, picture messages, and video messages are transmitted to the server. Such non-real-time transmission services are generally one-off best-effort transmission. The higher the rate requirement, the better the service experience. The service experience mainly depends on the transmission latency, which is closely related to the rate and file size and is not an objective indicator. The uplink rate is used as the core evaluation indicator.

Table 7-1 Core factors affecting the uplink real-time streaming media transmission services

Factor	Impact
Bit rate	The bit rate refers to the number of audio or video bits transmitted or processed per unit time. It is a common indicator for measuring the audio and video quality. Specifically, a high resolution, high frame rate, and low compression rate usually lead to an increase in bit rates given the same coding used.
Frame rate	The frame rate indicates how frequently pictures appear on display continuously in the unit of frame. The frame rate of the video content must be compatible with the frame rate attribute of the display device. For example, live broadcast services have higher requirements on frame rate stability. Frame rate fluctuation may deteriorate the quality of transmission videos in live broadcast.
Resolution	The video resolution indicates to the number of pixels contained in the video content. The video resolution must be compatible with the resolution of the display device. Otherwise, the video resolution may decrease or the video may not be displayed. For real-time live broadcast transmission services, the resolution is fixed and is closely related to the capabilities of terminals (such as cameras).
Packet loss	Packet loss has a significant impact on the quality of multimedia content. When a packet including an I-frame is lost, all subsequent frames of a same GOP frame depending on the frame are lost, which, as a result, could cause pixelization, frame blocking, and video output stalling. This can also be applied to audio streams. Packet loss can be measured by the average packet loss rate or burst packet loss rate. Burst packet loss has a greater impact on the system. Therefore, it must be considered separately.
Data packet latency	When a packet is transmitted from the source to the destination, transmission latency occurs. If the latency reaches a certain threshold, image blocking and image damage may occur.

Factor	Impact
Jitter	The propagation latency is not constant in a period of time. Therefore, the latency is changing across the entire network. Jitter is an indicator to measure this variability. The real-time streaming media service requires stable IP streams. Jitter may cause buffer overflow and underload, resulting in pixelization and frame freezing of the streaming content.
Average throughput	Rate is a key indicator for ensuring video transmission quality. The average rate alone is not enough. For real-time transmission services, rate fluctuation causes buffer overflow during video transmission. As a result, video frames cannot be played smoothly.
Peak throughput	Peak throughput in the real-time streaming media transmission service is measured by the bit rate when the quality is high. In fact, the peak throughput usually does not reflect the network transmission capability. Therefore, the peak throughput is the maximum throughput that can be reached instantaneously. This number reflects the transmission performance of the pipe.

Throughput swing	Throughput swing is defined as the proportion of the throughput that exceeds that of the previous or next session. It indicates the throughput fluctuation.
	
	The average performance cannot reflect the uplink multimedia transmission service experience. Burst congestion or deterioration will adversely impact user experience.
	● The "dynamic index" and "swing index" are introduced to reflect the fluctuation.
	  ● The proportion of the upward fluctuation that exceeds the range of μ+3σ is the upward swing index.
	  ● The proportion of the downward fluctuation that exceeds the range of μ-3σ is the downward swing index. It is the most essential indicator of quality deterioration.
	  ● For uplink real-time transmission services, a lower swing index indicates stabler transmission performance and better user experience.
	● From the perspective of real-time measurement, the values of μ and 3σ are calculated based on the average value and variance of the current time. Therefore, their values change dynamically. In the figure, μ and 3σ are represented as an f(x) curve.

Mobility interruption time	Mobility interruption time refers to the service interruption latency generated when a terminal moves. This indicator does not apply to fixed terminals.
Interactive latency	Interactive latency indicates the latency of the interactive behavior generated during user operations such as camera switch or video playing or pausing. This latency affects user experience in real-time operations.

The throughput stability is proposed in this paper, according to the research on the real-time streaming media protocol in ITU-T P.1201.

Table 7-2 Core factors affecting the uplink multimedia message transmission services

Factor	Impact
Transmission waiting time	Under poor network performance, the transmission waiting time is long, adversely affecting user experience. However, in this scenario, the transmission waiting time is closely related to the size of the file to be transmitted. Therefore, it cannot reflect the objective service quality, and is not recommended to be used for evaluation and monitoring.
Uplink throughput	Throughput is the key to guaranteeing the quality of video transmission. Multimedia message transmission is essentially a file uploading process. Throughput assurance is the key. Unlike real-time streaming media services, uploading services are best-effort services, which reflect the maximal uplink transmission performance of pipes.
Throughput swing index	In upload services, the file size varies according to the enterprise requirements. When a large file is transmitted, the transmission latency is high. In this case, the throughput fluctuation affects the waiting time and user experience.

7.1.2 Indicator System

Table 7-3 Indicator system of uplink real-time streaming media transmission services

Layer	Protocol	Indicator	Indicator Measurement Description
Comprehensive Score	E-Score		
Media quality index (MQI)	Media Quality Index		
	MPEG	Video resolution	Generally, the resolution of the camera is fixed.
	MPEG	Video bit rate	Generally, the average bit rate of the camera is fixed.
	MPEG	Video frame rate	Generally, the average frame rate of the camera is fixed.
	MPEG	Video encoding and decoding	H.264/H.265 image compression encoding
	MPEG	Audio bit rate	Generally, the average bit rate of the camera is fixed.
	MPEG	Audio frame rate	Generally, the average frame rate of the camera is fixed.
Interaction quality index (IQI)	Interaction Quality Index		
	SDK	Interactive latency	Latency from the time a control message is sent to the time the

Layer	Protocol	Indicator	Indicator Measurement Description
			message is responded.
	MPEG	Encoding latency	Encoding latency of the camera
	MPEG	Decoding latency	Decoding latency of the decoding server
	RTSP	Video playback latency	Latency from PLAY to 200OK
	RTSP	Video pause latency	Latency from PAUSE to 200OK
	RTCP	Round-trip latency	Calculated based on the RTCP timestamp
	RTCP	Latency jitter	Calculated based on the round-trip latency
Presentation quality index (PQI)	Presentation Quality Index		
	SDK	Slice	Obtaining the decoding server SDK
	SDK	Stall	Obtaining the decoding server SDK
	RTP	Average uplink throughput	Calculating the average value at the flow level
	RTP	Uplink peak throughput	Setting the sampling window and maximum measurement value
	RTP	Uplink throughput swing index	Fluctuation of the experience throughput
	RTP	Average transmit packet discard rate	Calculated based on the RTP sequence
	RTP	Burst transmit packet discard rate	Calculated based on the RTP sequence
	RTCP	Round-trip latency	Calculated based on the RTCP timestamp
	RTCP	Latency jitter	Calculated based on the round-trip latency

In the uplink multimedia message transmission scenario:

Table 7-4 Indicator system of uplink multimedia message services

Category	Indicator
Comprehensive service quality evaluation	Transmission quality index/E-Score
Uplink transmission quality	UL Average Throughput
	UL Peak Throughput
	UL Throughput Swing Index

7.1.3 Modeling Method

Figure 7-1 E-Score modeling framework for uplink real-time streaming media transmission services

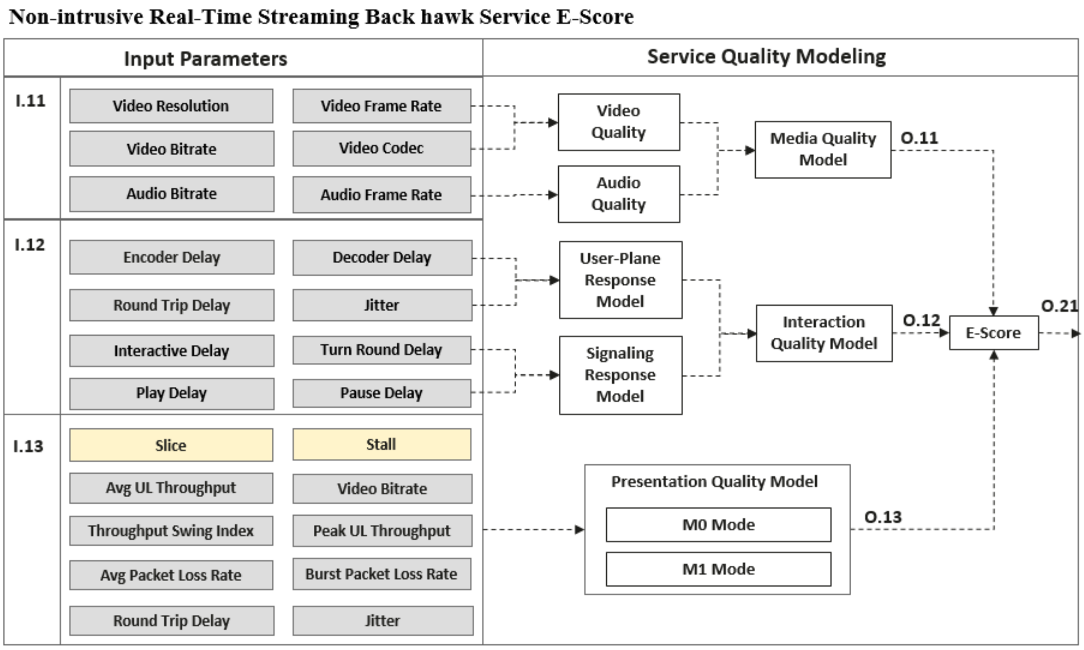


Table 7-5 I.11–input parameters of media quality

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Video bit rate	$VideoBr$	Float, kbps	Per segment	Mode 0	MQI
Video frame rate	$VideoFr$	Integer	Per segment		
Measurement interval	TPD	Float, ms	Per segment		

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Video resolution	Res	Length x width 2880 x 1600	Per segment		
Video encoding	$VideoCodec$	One of: H264-baseline, H264-high, H264-main, H265-high, H265-main	One Session		
Object moving speed	*objSpeed*	Float, Km/s	Per segment		
Object distance	*obDistance*	Float, m	Per segment		
Screen size	*screenSize*	Float, in	One Session		
Audio bit rate	$AudioBr$	Float, kbps	One Session		
Audio encoding	$AudioCodec$	Integer	One Session		

Table 7-6 I.12–Input parameters of interaction quality

Input Parameter Name	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Encoding latency	D_e	Float, ms	Per Segment	Mode 0	IQI
Decoding latency	D_d	Float, ms	Per Segment		
Interactive latency	D_i	Float, ms	Per Segment		
Round-trip latency	D_{rt}	Float, ms	Per Segment	Mode 0 Mode 1	
Latency jitter	$jitter$	Float, ms	Per Segment		
Playback latency	$PlayDelay$	Float, ms	One Session		
Pause latency	$PauseDelay$	Float, ms	One Session		

Table 7-7 I.13–Input parameters of presentation quality

Input Parameter Description	Abbreviation	Value	Obtaining Frequency	Data Source	Module
Pixelization start time	$Slice_{BT}$	Float, ms	Per Segment	Mode 0 Mode 1	PQI
Pixelization end time	$Slice_{ET}$	Float, ms	Per Segment		
Frame freezing start time	$Stall_{BT}$	Float, ms	Per Segment		
Frame freezing end time	$Stall_{ET}$	Float, ms	Per Segment		
Video bit rate	$VideoCodec$	Float, kbps	One Session		
Average throughput	Thr_{avg}	Float, kbps	Per Segment		
Peak throughput	Thr_{max}	Float, kbps	Per Segment		
Throughput upward swing index	Thr_{ulsi}	Float, %	Per Segment		
Throughput downward swing index	Thr_{dlsi}	Float, %	Per Segment		
Average packet loss rate	ppl	Float, ms	Per Segment		
Burst packet loss rate	bpl	Float, %	Per Segment		
Latency jitter	$jitter$	Float, ms	Per Segment		

- The formula for calculating the uplink real-time streaming media service indicators is as follows:

2. Media quality (enhanced based on ITU-T P.1201)

[E-Score]

$$EScore = \omega_1 * MQI + \omega_2 * IQI + \omega_3 * PQI$$

[Media quality index]

$$MQI = c1 * QA + c2 * QV$$

[Audio quality]

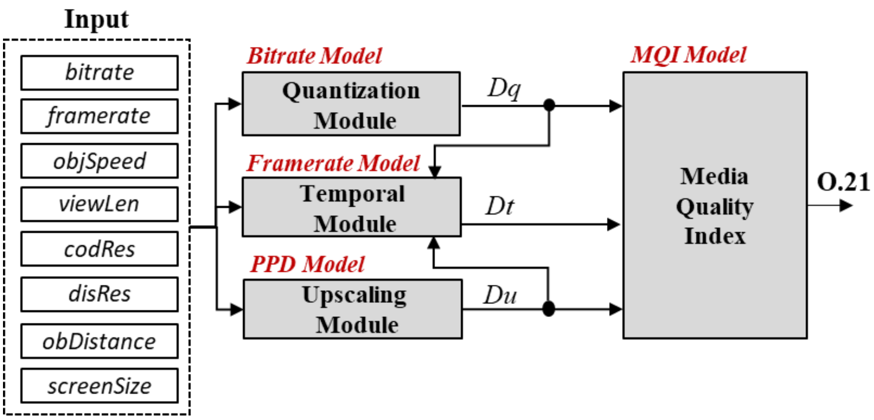
$$QA = I_A - QcodA$$

QA is the predicted audio quality. $QcodA$ is the quality impairment caused by audio compression. I_A is the impact of different audio encoding formats $AudioCodec$. The score of the PCM encoding format is 100, and other audio encoding formats (AAC-LC, HE-AACv2, MPEG1-LII, AC3) all have quality impairment.

$$QcodA = a_{1A} * \exp(a_{2A} * AudioBr) + a_{3A}$$

[Video quality]

Figure 7-2 Modeling framework for video quality index



- Dq (Quantization Degradation): impact of a unit quantity of quantized bits on video quality
- Dt (Temporal Degradation): impact of time complexity on video quality
- Du (Upscaling Degradation): impact of spatial complexity on video quality

$$QV = f\left(D_q, D_u, D_t\right)$$

$$D_q = f_1(disRes, bitrate, sceneType)$$

$$D_u = f_2(disRes, codRes, obDistance, screenSize)$$

$$D_t = f_3(framerate, objSpeed, viewLen, D_q, D_u)$$

[Presentation quality index]

[Mode0]

$$PQI = b_{1V} * \log((b_{2V} * FreezingRatio + b_{3V} * SlicingRatio) + 1)$$

Of which:

$$FreezingRatio = \frac{FreezingDuration}{TotalDownloadDuration}$$

$$SlicingRatio = \frac{SlicingDuration}{TotalDownloadDuration}$$

[Mode1]

$$PQI = c_{1V} * \log(\left(c_{2V} * Q_{bandwidth} + c_{3V} * Q_{packetloss} + c_{4V} * Q_{jitter}\right) + 1)$$

$$Q_{bandwidth} = b_1 * \frac{Thr}{VideoBr} * \exp(b_2 * ULSI + b_3 * DLSI + b_4)$$

$$Q_{packetloss} = p_1 \times \frac{ppl}{\frac{ppl}{BurstR} + bpl} + p_0$$

$$Q_{jitter} = j_1 * \exp(j_2 * Jitter + j_3)$$

In the formula, $BurstR$ indicates the impact factor of burst packet loss on services compared with random packet loss. For live video, the value is fixed.

[Interaction quality index]

$$IQI = 100 - Q_{up} - Q_{sp}$$

[User-plane response model]

$$Q_{up} = a_{1I} * log(a_{2I} * (D_{rt} + D_e + D_d) + a_{3I} * Jitter + a_{4I})$$

[Control-plane response model]

$$Q_{sp} = b_{1I} * log(b_{2I} * (D_i) + b_{3I})$$

In the formula, D_i can be obtained in different methods. Some applications may directly obtain the value from a control-plane interaction message, and others may obtain the value from a protocol. For a play action in the RTSP protocol, $D_i = PlayDelay$, and for a pause action, $D_i = PauseDelay$. Success rate indicators are not closely related to the network. Therefore, you are advised not to include them into the calculation of interaction quality.

- The formula for calculating the uplink multi-media message service is as follows:

For this best-effort multimedia file uploading behavior, consider the uplink bandwidth and its stability.

$$QM = 100 - I_{bandwidth}$$

$$I_{bandwidth} = \begin{cases} 0, & Thr < Thr0 \\ (b_1 \times Thr + b_0) \times \exp(b_2 \times ULSI + b_3 \times DLSI + b_4), & Thr \geq Thr0 \end{cases}$$

NOTE

3. There is no technical difficulty in measuring the throughput and the data is easy to obtain. However, you might run into technical difficulties when trying to measure the throughput fluctuation, as it depends on products or small-granularity statistics for implementation (e.g. probe-based dotting, such as MR data implementation). Additionally, the throughput fluctuation measurement must be performed in a specific time window, because indicator value varies significantly in different time windows.
4. Latency measurement is also technically difficult, as in most scenarios, it is difficult to obtain precise latency metrics. This document provides an end-to-end latency model, which is a non-intrusive small-granularity latency evaluation algorithm. This algorithm can be used to obtain latency indicators, and can break down the latency problem to radio factors such as congestion, interference, and coverage, and provide the requirement boundary of each factor. In this way, the problem can be demarcated for subsequent optimization.

7.1.4 Experience Baseline

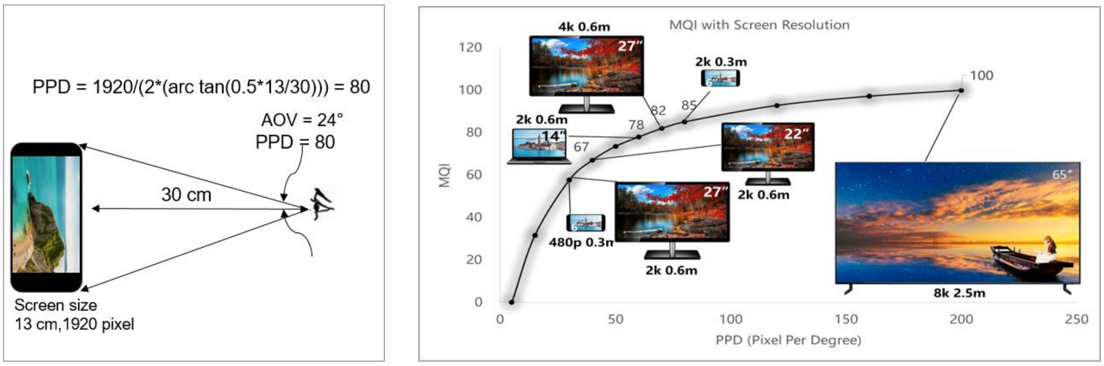
7.1.4.1 Remote Video Control – Video Surveillance Experience Baseline

The remote video control service means that a remote operator controls a device by using a dedicated control system according to real-time video surveillance images, to complete numerous service operations.

Experience modeling refers to analysing video surveillance of the remote video control service at site X to obtain the experience baseline.

[PPD model] Impact of pixel per degree (PPD) on video experience

Figure 7-3 Impact of PPD on video experience



Screen size, viewing distance, and field of view (FoV) all affect video experience and can be subsequently converted into PPD for measurement. The curve in the preceding figure shows an association between PPD and video experience, and once PPD increases to a certain value, the MQI improvement becomes smaller. The model analysis is based on the video surveillance scenarios of Hunan Valin Xiangtan Iron & Steel Co., Ltd, which has a typical 27-inch screen configuration. The corresponding FoV is 53° and the viewing distance is 60 cm.

Table 7-8 Mapping between PPD and MQI values

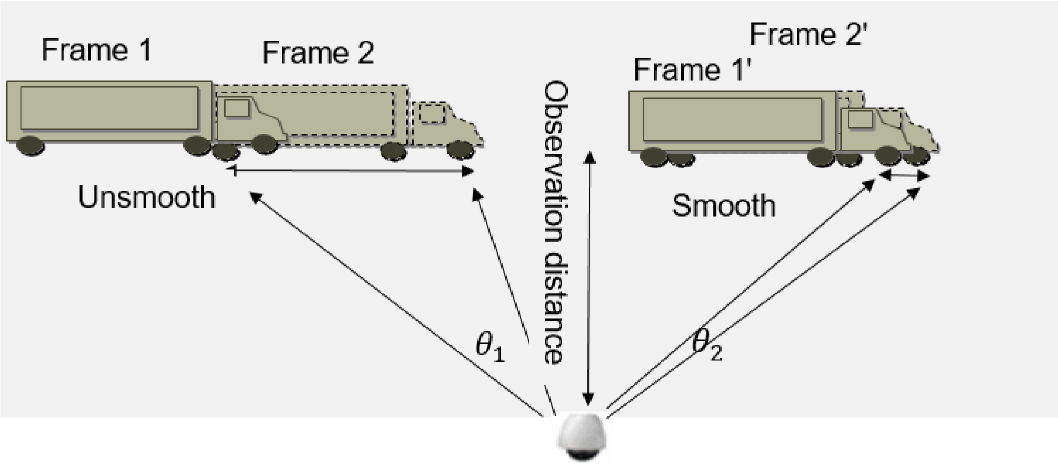
Screen Size (Inch)	Resolution	Display Mode	PPD	Equivalent Mobile Phone Resolution	MQI
27	3840 x 2160	Main display	70	Approx. 1080p	82
27	2560 x 1440	Main display	48	> 720p	73
27	1920 x 1080	Main display	36	Approx. 720p	63
27	1920 x 1080	Auxiliary 4-split screen	70	Approx. 1080p	82
27	1080 x 720	Auxiliary 4-split screen	41	> 720p	67
27	720 x 480	Auxiliary 4-split screen	27	Approx. 600p	51

Key conclusions:

1. PPD is a key factor that affects video media quality from the spatial information dimension, comprising multiple factors such as resolution, screen size, and viewing distance.
2. Good experience (80+ points): 70 PPD (27-inch screen, 4K resolution, 0.6 m viewing distance); fair experience (60+ points): 35 PPD (27-inch screen, 2K resolution, 0.6 m viewing distance)
3. Recommended surveillance screen resolution = Maximum camera resolution

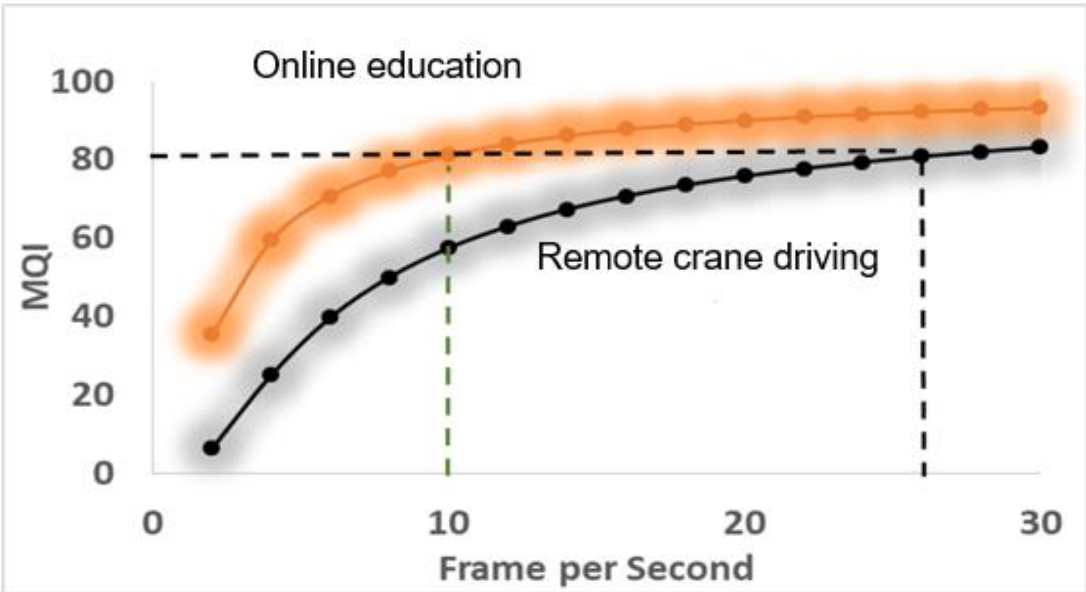
[FPS model] Frames per second (FPS) impact on video experience

Figure 7-4 FPS required by object movement and observation distance



In the remote video control service, images are constantly moving, and the speed and viewing distance of the images greatly affect video quality.

Figure 7-5 FPS impact on video experience of different services

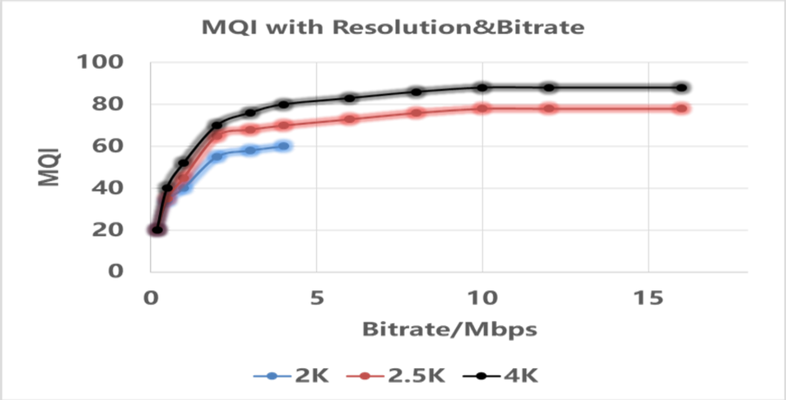


Key conclusions:

1. A faster image speed and shorter distance require a higher FPS, whereas a slower image speed and longer distance require a lower FPS.
2. The longer the distance, the lower the object definition. In industrial scenarios, the distance is determined by the industrial design. Within the design scope, the camera should be as close to the observed object as possible.
3. In a remote video control scenario (ViewLen: 9 m; ObjMoveSpeed: 1.2 m/s), the FPS for good experience (80+ points) is 25 and for fair experience (60+ points) is 12.

[Bitrate model] Impact of the bitrate on video experience

Figure 7-6 Impact of bitrate changes in different resolutions on video experience



In low-speed industrial automation scenarios (5 km/h), the resolution has a greater impact on user experience than the bitrate. As shown in the preceding figure, the improvement in experience is not obvious when the bitrate is higher than 2 Mbps, 10 Mbps, and 4 Mbps in 2K resolution, 2.5K resolution, and 4K resolution, respectively.

Table 7-9 Mapping between bitrates and MQIs in different resolutions

Coding Standard	Resolution	Bitrate	MQI
H.265	2K	2 Mbps	56
H.265	2K	4 Mbps	60
H.265	2.5K	2 Mbps	65
H.265	2.5K	4 Mbps	70
H.265	2.5K	10 Mbps	78
H.265	4K	2 Mbps	70
H.265	4K	4 Mbps	80
H.265	4K	10 Mbps	88

Key findings:

1. In low-speed industrial automation scenarios (5 km/h), space complexity contributes greater to image quality, while the resolution has a significant impact on user experience compared with the bitrate.
2. Good experience in industrial automation low-speed scenarios (5 km/h): When the video resolution is 2.5K/4K, the recommended bitrate is 10 Mbps/4 Mbps.
3. Fair experience in industrial automation low-speed scenarios (5 km/h): When the video resolution is 2K/2.5K/4K, the recommended bitrate is 4 Mbps/2 Mbps/2 Mbps.

[iTBR model] Impact of I frame throughput-to-bitrate ratio (iTBR) on video experience

The transmission rate of the video surveillance service differs from that of the video on demand (VoD) service, and refers to the I-frame transmission rate rather than the average transmission rate.

Figure 7-7 Relationship between the iTBR and video experience

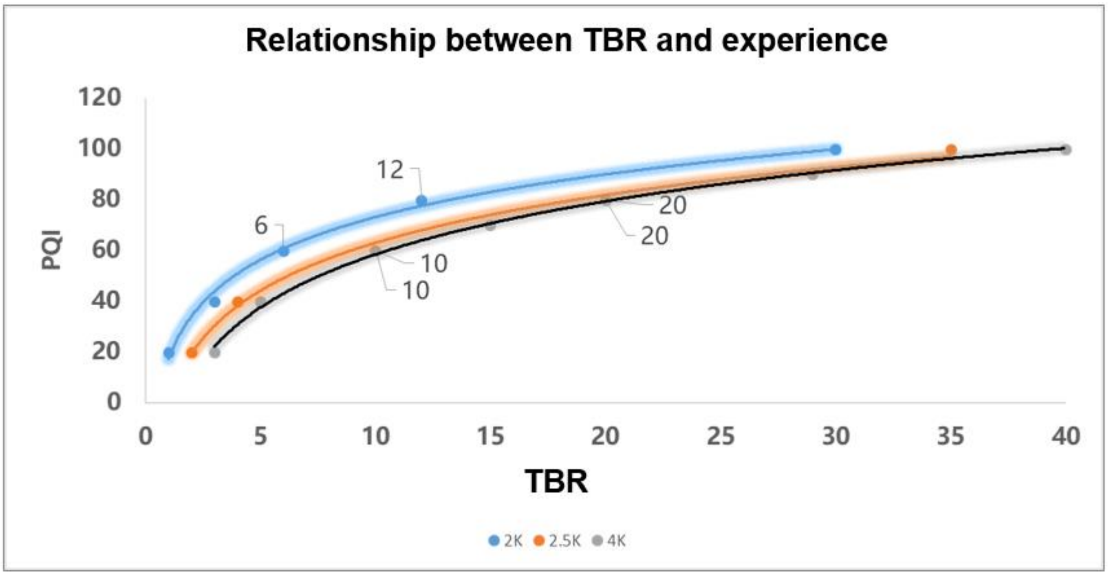


Table 7-10 Experience baseline of the iTBR

Resolution	I-Frame Size (Mbit)	I-Frame Rate (Mbps)	Bitrate (Mbps)	TBR (Fair Experience)	TBR (Good Experience)
4K	3.4	34	4	10	20
4K	9.1	91	8		
4K	14.1	141	16		
2.5K	2.2	22	2	10	20
2.5K	4.2	42	4		
2.5K	7.6	76	8		
2K	1.1	11	2	6	12
2K	2.3	23	4		

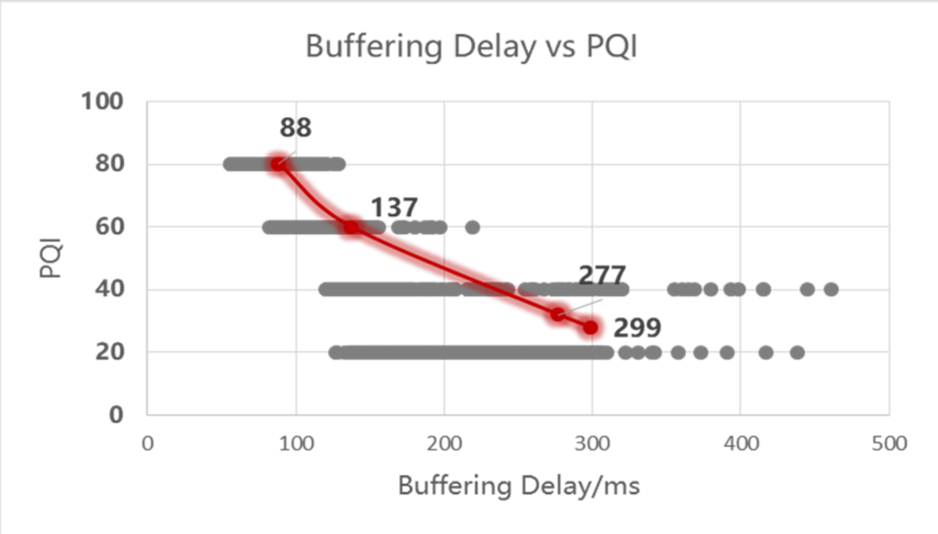
Key conclusions:

1. The I-frame buffer duration is a key factor that affects user experience and is related to the I-frame size and bitrate. With the same resolution and bitrate, an iTBR model can be established to ensure user experience.
2. To ensure a good experience (80+ points), the recommended iTBR for 2K/2.5K/4K is 12/20/20, respectively. To ensure a fair experience (60+ points), the recommended iTBR for 2K/2.5K/4K is 6/10/10, respectively.

[Burst model] Impact of the I-frame buffering delay on video experience

The buffering delay refers to the interval between when a frame is successfully displayed and when the next I-frame is displayed. It ensures that video images are continuously displayed.

Figure 7-8 Relationship between the buffering delay and video experience



Key conclusions:

In a typical 4K, 4 Mbps bitrate, and 25 FPS (frame interval: 40 ms) scenario:

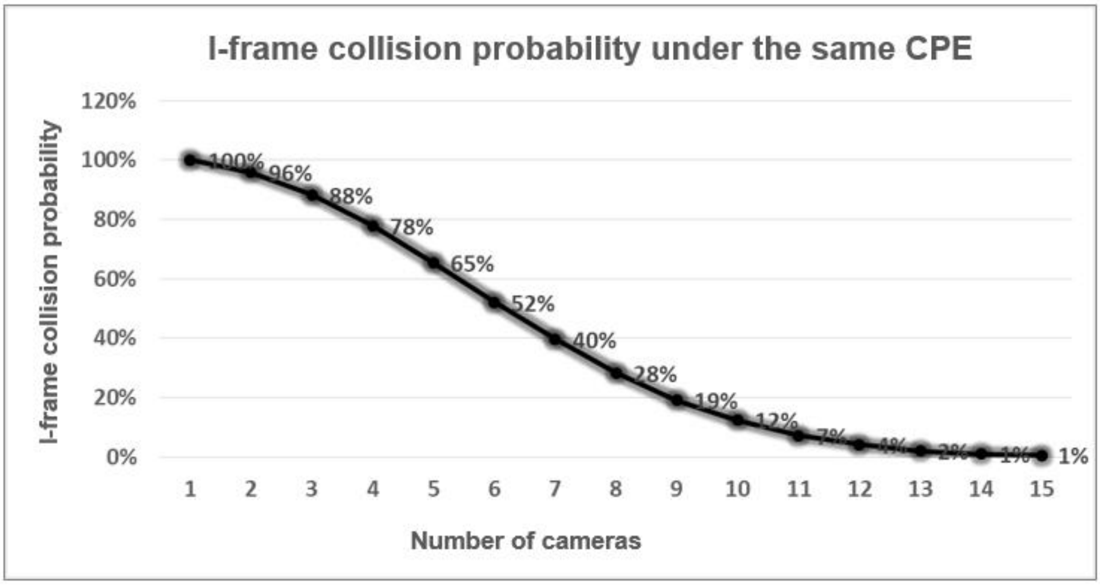
1. Good experience (80+ points): When the buffering delay is less than or equal to 90 ms, videos run smoothly.
2. Fair experience (60+ points): When the buffering delay is less than or equal to 140 ms, some viewers may notice frame freezing.

[Multi-channel collision model] Multi-camera peak bandwidth collision probability

$$P(ColN, CamN, IFI, IFL) = \frac{C_{CamN}^{ColN} * \left(\frac{IFI}{IFL} - 1\right)^{CamN-ColN} * C_{\frac{IFI}{IFL}}^{1}}{IFI^{CamN}}$$

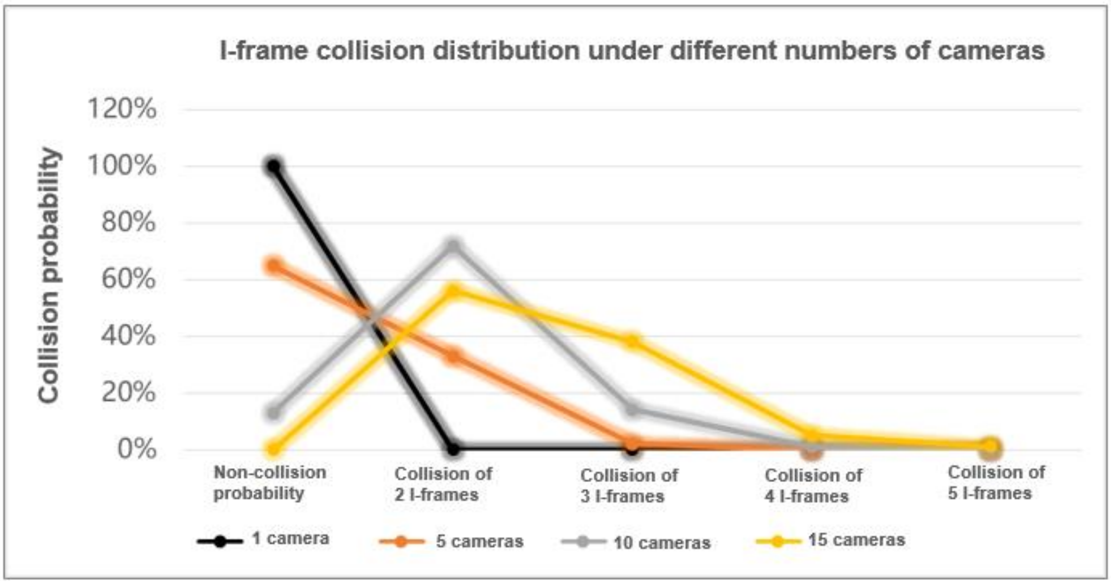
ColN indicates the number of collision frames, **CamN** indicates the total number of cameras, **IFI** indicates the I-frame interval, and **IFL** indicates the transmission duration in frames.

Figure 7-9 Relationship between the number of cameras and the I-frame collision probability



- When cameras start randomly, the more cameras there are, the higher the I-frame collision probability.
- With six cameras, the I-frame collision probability is 50%.
- With 11 cameras, the I-frame collision probability reaches 93%.

Figure 7-10 Probability analysis of concurrent I-frame collisions



The number and probability of I-frame collisions under multiple deployed cameras are as follows:

- With five cameras, the collision probability of three I frames is 2%.
- With 15 cameras, the collision probability of four I frames is 5%.

Table 7-11 Number of I-frame collisions and occurrence probability when multiple cameras are used concurrently

Number of Cameras	Non-collision Probability	Collision of Two I Frames	Collision of Three I Frames	Collision of Four I Frames	Collision of Five I Frames	Collision of Six I Frames	Collision of Seven I Frames	Collision of Eight I Frames	Collision of Nine I Frames
1	100%	0%	0%	0%	0%	0%	0%	0%	0%
5	65%	33%	2%	0%	0%	0%	0%	0%	0%
10	13%	72%	14%	1%	0%	0%	0%	0%	0%
15	0%	56%	38%	5%	1%	0%	0%	0%	0%
20	0%	25%	57%	15.6%	2.2%	0.02%	0.02%	0%	0%
30	0%	0.9%	40%	43.9%	12.6%	2.3%	0.3%	0%	0%
40	0%	0%	8%	47.8%	32.4%	9.5%	2%	0.3%	0%
50	0%	0%	0.3%	22%	45.6%	23.3%	6.8%	1.6%	0.3%

To ensure that the non-frame-freezing rate is greater than 99%, the bandwidth required by multiple services is calculated as follows:

Bandwidth required by multiple services = Bitrate x I-frame bandwidth multiple x Number of collision channels + Bitrate x Number of non-collision channels

For a 25 FPS and I-frame interval of 50-cycle scenario, the collision probability can be reduced by configuring different I-frame intervals for the cameras. The following uses 4 megapixels and 2 Mbps bitrate as an example:

When there are 15 cameras with a 1% probability of collisions for five I-frames, the bandwidth must be four times bigger than the I-frame bitrate.

I-frame bitrate = 2 x 10

Experience baseline = 2 x 10 x 4 + 2 x 11 = 102 Mbps

The network capability determines the type of services that can be carried on a network. However, the FPS, bitrate, resolution, and codec in video surveillance scenarios can be adjusted based on service requirements and network capabilities.

Table 7-12 Experience baseline of the video surveillance service

Typical Application Instance	Impact Factor		Network Capability for Ensuring Fair Experience			Network Capability for Ensuring Good Experience		
	Typical Bitrate	**Number of Cameras**	**One-way Network Latency**	**Required Bandwidth**	**TBR**	**One-way Network Latency**	**Required Bandwidth**	**TBR**
Video surveillance (2K)	2 Mbps/H.265	5	< 25 ms	> 30 Mbps	> 6	N/A*		
	2 Mbps/H.265	10	< 25 ms	> 40 Mbps	> 6			
	2 Mbps/H.265	20	< 25 ms	> 70 Mbps	> 6			
Video surveillance (2.5K)	2 Mbps/H.265	5	< 25 ms	> 46 Mbps	> 10			
	2 Mbps/H.265	10	< 25 ms	> 74 Mbps	> 10			
	10 Mbps/H.265	5	< 25 ms	> 130 Mbps	> 10	< 10 ms	> 430 Mbps	> 20
	10 Mbps/H.265	10	< 25 ms	> 370 Mbps	> 10	< 10 ms	> 670 Mbps	> 20
Video surveillance (4K)	4 Mbps/H.265	5	< 25 ms	> 92 Mbps	> 10	< 10 ms	> 87 Mbps	> 20
	4 Mbps/H.265	10	< 25 ms	> 148 Mbps	> 10	< 10 ms	> 268 Mbps	> 20

7.2 Downlink Multimedia Transmission Service

7.2.1 Impact Factor

The downlink streaming transmission service described in this section is the same as the traditional B2C streaming transmission service, where the only difference is the carrier. The downlink streaming transmission service is classified into the following types:

- Video on-demand/live streaming: for example, in the in-car entertainment scenarios of vehicle-to-everything (V2X)
- Cloud VR: ToB-oriented service applications, such as VR education scenarios of distance education
- Cloud PC: ToB-oriented service applications, such as cloud office

The traditional video streaming media evaluation system is mature. For details, consult white papers covering service experience standards such as HD video, cloud VR, and cloud PC.

7.2.2 Indicator System

Table 7-13 Indicator system of downlink streaming services

Category	Indicator
Comprehensive service quality evaluation	**Q-score (backhaul)**
Media quality index (MQI)	Video bitrate
	Video frame rate
	Resolution
Presentation quality index (PQI)	Average packet loss rate
	Burst packet loss rate
	Round trip delay
	Delay tail index
	Jitter
	UL average throughput
	UL peak throughput
	UL throughput swing index
	Mobility interruption time
Interaction quality index (IQI)	Round trip delay
	Encoder delay
	Decoder delay
	Rendering delay

7.2.3 Modeling Method

The traditional video streaming media evaluation system is mature. For details, consult white papers covering service experience standards such as HD video, cloud VR, and cloud PC.

7.2.4 Experience Baseline

Dependent on project requirements.

7.3 AR Service

7.3.1 Impact Factor

This section analyzes the fast-growing augmented reality (AR) services in ToB scenarios.

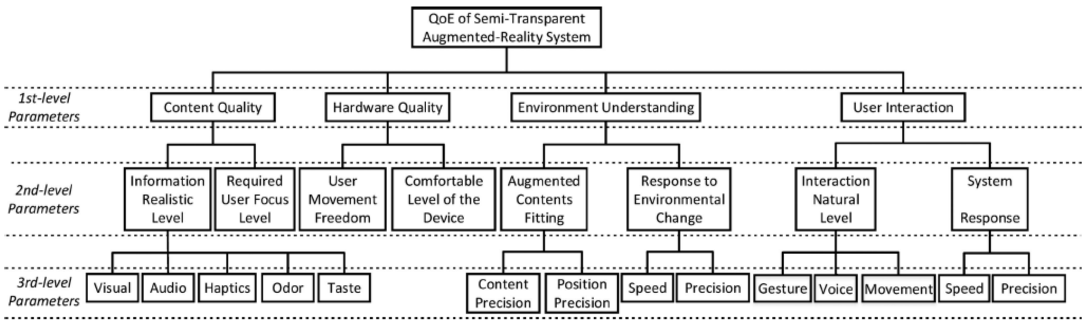
AR+5G is an AR-based remote video visualization solution. It allows real-time HD live interaction with back-end experts through the 5G network, without requiring manual operation. This solution enables experts to provide online and remote guidance at any time, and is applicable to multiple industries including communications, medical care, manufacturing.

The AR applications dedicated for AR smart glasses (ARSGs), such as VR glass and Microsoft Hololens, offer an innovative way for users to interact with AR content. ARSGs have attracted the attention of the medicine, tourism, education, and manufacturing fields.

To evaluate AR service experience, we propose a QoE framework, which consists of three impact parameter levels. This framework uses a fuzzy inference system modeling method to quantitatively evaluate user experience of ARSGs.

The first level of the framework covers four aspects: content quality, hardware quality, environment understanding, and user interaction.

Figure 7-11 Factors affecting AR services



The following table describes the core factors.

Table 7-14 Core factors affecting AR services

First Level	Second Level	Third Level	Impact
Content quality	Information realistic level	Visual	Major factor, affecting image definition
		Audio	Major factor, affecting voice definition
		Haptics	Minor factor
		Odor	Minor factor
		Taste	Minor factor
	Required user focus level	Focus level	Level of attention users must spend on AR content, as required by an application. A higher level indicates a higher requirement on the interaction latency and AR content quality.
Hardware quality	User movement freedom	DOF	Freedom of interaction between users and AR content.
	Device comfort level	Device comfort level	Comfort degree of wearing ARSGs or head-mounted displays (HMDs)
Environment understanding	Augmented contents	Content precision	Matches the AR content within the external environment. For example, place a cup rather

First Level	Second Level	Third Level	Impact
	fitting		than a car on a table.
		Position precision	Matches the positioning of AR content within the external environment. For example, place a cup on the table rather than letting it float around.
	Response to environmental change	Speed	Latency in response to AR content and environmental changes.
		Precision	Positioning precision in response to AR content and environmental changes.
User interaction	Interaction natural level	Gesture	Gesture interaction between users and AR content, and ease of use. For example, click a specified button to bring an AR object closer.
		Voice	Voice interaction between users and AR content, and easier usability. For example, use a voice instruction to move an AR object closer to the user.
		Movement	Mobile interaction between users and AR content, and easier usability. For example, use a body motion to move an AR object.
	System response	Speed	Interaction latency between users and AR content.
		Precision	Represents whether the interaction result between a user and AR content is correct or incorrect.
Presentation experience (added dimension for comprehensively measuring user experience)	User experience	Load latency	Refers to the latency a user experiences when accessing an AR application. In most cases, latency is determined by how quickly content data loads. In addition to network performance, cloud and terminal encoding/decoding as well as rendering latency affect the loading latency.
		Frame freezing	Frame freezing occurs in AR HD videos, due to bandwidth availability.
		Blurring	Blurry AR content may be prevalent due to scenario construction quality or network packet loss.
		Background movement	The background may move during interaction when the AR content does not match the environment
		Visual clutter	Content may appear cluttered when the AR content does not match the environment.

1. Among AR service modeling elements, the highlighted network factors in the preceding table play a major role.
2. The commercial use of AR services may involve video call behaviors, which need to be analyzed separately from AR behaviors.

7.3.2 Indicator System

Table 7-15 AR service indicator system

Category	Subcategory	Indicator
Q-score (AR)		
Media quality index (MQI)	Hardware quality	Screen frame rate
		Screen resolution
	Video quality	Video bitrate
		Video frame rate
		Resolution
	Audio quality	Audio bitrate
		Audio frame rate
	Audio-video synchronization	A-V synchronization
Interaction quality index (IQI)	Operation experience	Degree of freedom (DOF)
		Content precision
		Space precision
		Level of usability
	Response experience	Environment response delay
		Environment response spatial precision
		Content response delay
		Content response spatial precision
		Encoder delay
		Decoder delay
		Rendering delay
		Round trip delay
		Delay tail index
		Jitter
Presentation quality index	Presentation experience	Loading delay

Category	Subcategory	Indicator
(PQI)		Blurriness
		Stalling
		Background movement
		Visual clutter
	Transmission quality	Average download throughput
		Peak download throughput
		Download throughput swing index
		Max. burst size
		Burst pulse number
		Burst packet loss rate

7.3.3 Modeling Method

Figure 7-12 AR service quality measurement framework

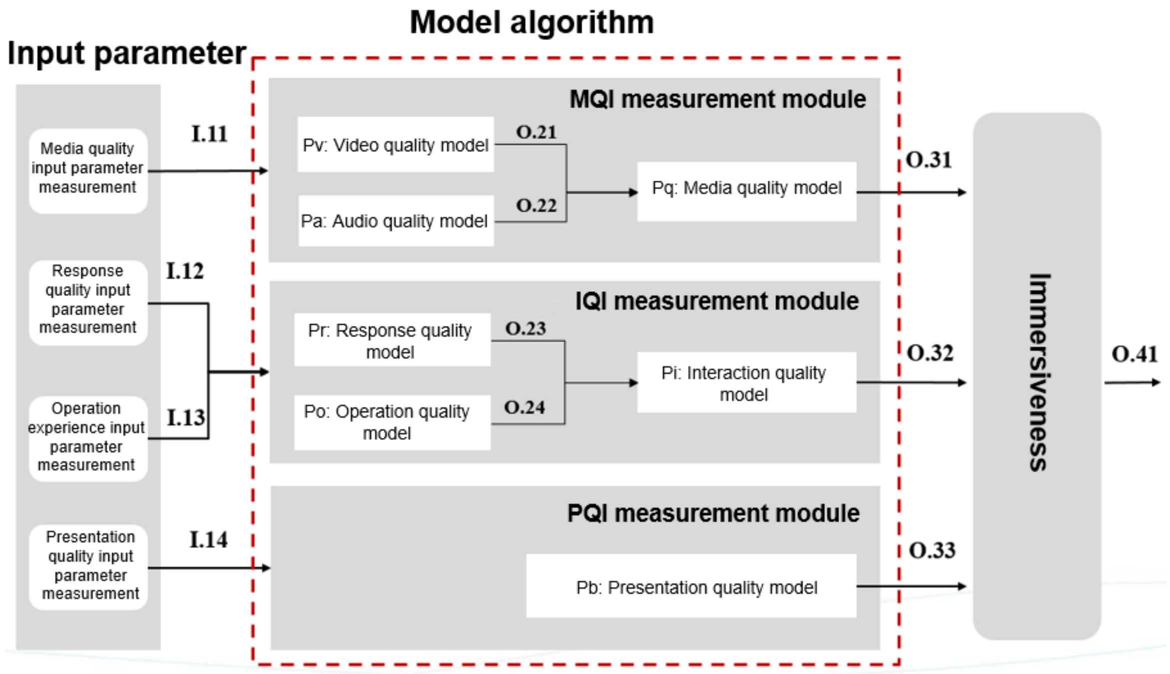
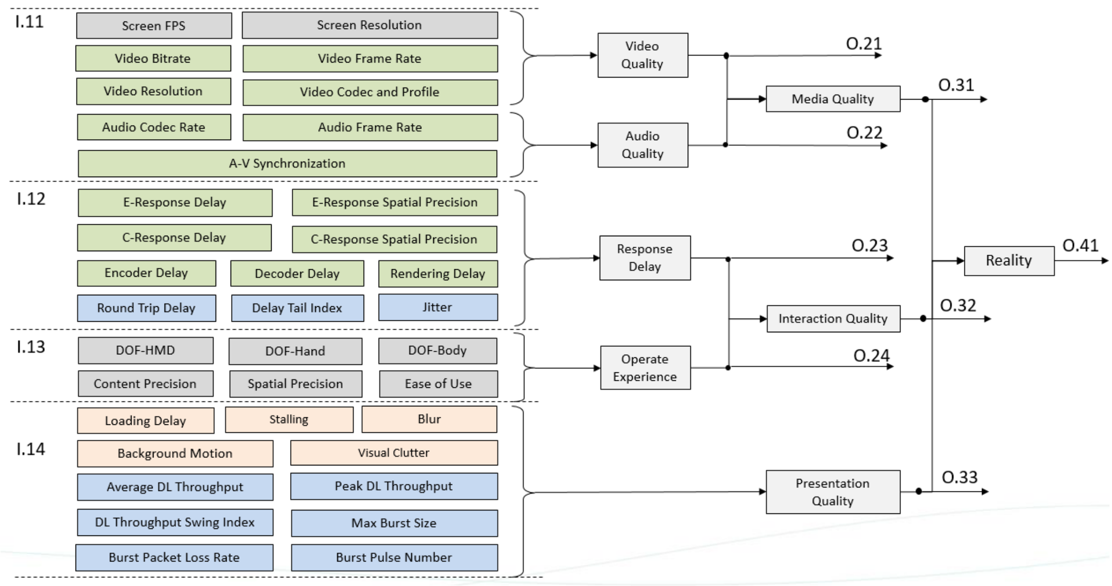


Figure 7-13 Input and output of AR service quality measurement



7.3.4 Experience Baseline

Dependent on project requirements.

7.4 Real-Time Interaction Service

7.4.1 Impact Factor

Real-time interaction services in ToB scenarios mainly refer to remote control and industrial control services. URLLC slicing is used in 5G SA networking to ensure ultra-low latency.

Real-time interaction services are mainly small-packet transmission services, which have no requirement for network bandwidth but require ultra-low latency. For industrial-grade URLLC services, high latency may not only lead to a deteriorated performance, but also potential service failures. Therefore, it is critical to analyze the boundary of latency. To prevent possible service failures, the network only needs to ensure that the latency is below the boundary. For example, if the SLA requires a 10 ms delay, the crucial point is to ensure that the delay of most service records is less than 10 ms or the number of service records whose delay exceeds the boundary is the minimum.

Table 7-16 Core factors that affect real-time interaction services

Factor	Impact
Interactive delay	It refers to the end-to-end delay from the time when the application layer requests a packet to the time when it responds to the packet.
Long tail of delay	Delay distribution is important for evaluating the overall service quality, especially the distribution of the tail of delay. In most cases, the tail of delay is prevalent in delay-sensitive services. The data distribution of the long tail is the key to meeting the delay boundary requirements.

Factor	Impact
Reliability	Reliability is a key factor that affects URLLC services. It can be divided into mean time between failures (MTBF), mean down time (MDT), and mean up time (MUT).
Availability	Availability indicates the probability of long-term running of a channel. The availability of a stable channel can also be interpreted as the ratio of the average channel running time.
Mobility interruption time	For URLLC services, the mobility interruption time must be 0.

Reliability and availability are system-level indicators. They are measured during long-term running and are not applicable to short-term service quality evaluation.

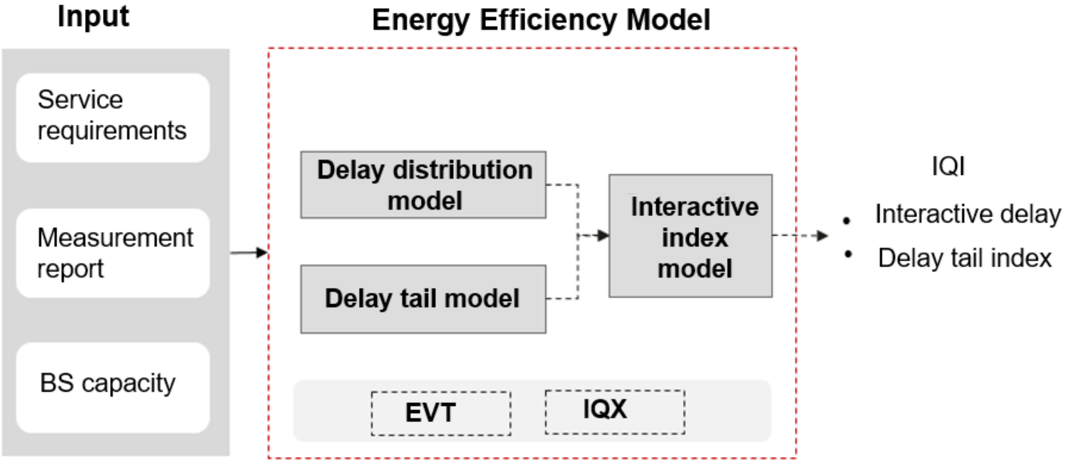
7.4.2 Indicator System

Table 7-17 Indicator system of real-time interaction services

Category	Indicator
Comprehensive service quality evaluation	**Interaction quality index (Q-Score)**
Service interaction quality (IQI)	Interactive Delay
	Interactive Delay Tail Index
	Mobility interruption time

7.4.3 Modeling Method

Figure 7-14 Interaction quality measurement method



1. It is difficult to measure the delay in terminal device content or in network images, so the latency measurement and collection methods need to be analyzed first.
2. Since the long tail distribution that exceeds the delay boundary greatly impacts services, the long tail phenomenon also needs to be evaluated.
3. The delay and tail indicators are both QoS indicators, to the relationship between QoS and QoE needs to be quantitatively described.

According to Weber's law, sensory differences can be perceived only when a physical stimulus changes more than the constant proportion of its actual stimulus. Fechner extended this basic relationship by assuming that the differential perception (dP) is proportional to the relative change dS/S of the human physical stimulus. That is, the relationship between the differential perception and the relative change of the stimulus for the Weber-Fechner Law is as follows:

$$dp = \frac{dS}{S}$$

Furthermore,

$$P = k \cdot ln\frac{S}{S_0}$$

Among which *P* indicates a perception amplitude and S_0 indicates a stimulus threshold.

According to the Weber-Fechner Law, the QoE-QoS mapping can be expressed as follows:

$$dQoS \propto QoS \cdot dQoE$$

Based on the IQX hypothesis theory, the following function is established:

$$I_{delay} = \begin{cases} 0, & MIT > 0 \\ d1 \cdot \exp(d_2 \times delay + d_3 \cdot DTI) + d4, & MIT = 0 \end{cases}$$

The values of *d1*, *d2*, *d3*, and *d4* can be obtained based on test data.

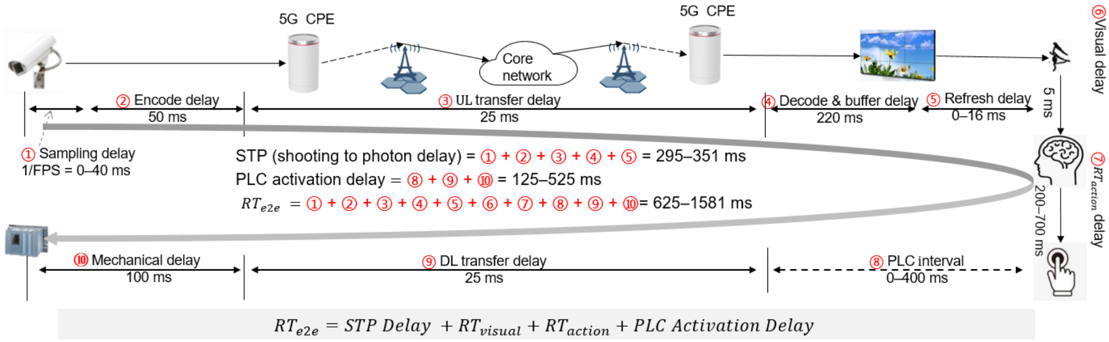
URLLC requires that the mobility interruption time (MIT) is 0, whereas the practical industrial system defines interruption as: the delay obtained through measurement or modeling in a movement process that is lower than the delay threshold. For example, a switchover does not cause service interruption or cause the delay to exceed the boundary. If the preceding conditions are met, the MIT is 0; otherwise, the MIT is not 0. If the MIT is not 0, the interaction index needs to be set to the minimum value.

7.4.4 Experience Baseline

The delay baseline for real-time interaction services varies depending on the projects. This section describes the experience baselines obtained based on the PLC scenario analysis of the remote video control service at site X.

[Control precision model] Relationship between control precision, movement speed, and delay

Figure 7-15 Logical decomposition of the E2E operation response delay



- **RT_e2e** (E2E reaction time) indicates the end-to-end operation response time. **[RT]_visual** indicates the visual signal response time. **RT_action** indicates the interval between the time when a signal is received and the time when a response action is taken.
- **STP_Delay** (shooting to photo delay) indicates the interval between the time when an image is shot by a camera and the time when the server displays the image.
- **PLC Activation Delay** indicates the interval between the time when a PLC signal is sent and the time when the device generates a response. **PLC Interval** indicates the interval for sending PLC signals.

Key findings:

1. If the delay of end-to-end operation responses is approximately 1s, the delay caused by device factors (codec, PLC, and machinery) is approximately 600 ms, the response time of personnel is approximately 200 ms, and the network delay is less than 50 ms.
2. The delay of end-to-end operation responses causes a control precision error of more than 1m. In the case of low speed operation, the control precision error can be reduced based on manual prediction.
3. If the delay is stable, the control precision error is easy to predict and the impact is small. If, however, the delay is unstable, the impact of burst traffic is unpredictable and can be significant. Therefore, the stability of delay must be ensured.

Summary:

The network capability determines the type of services that can be carried on a network. However, the movement speed, control period, and escape delay in the PLC scenario can be adjusted based on service requirements and network capabilities.

Table 7-18 Remote video control service – PLC experience baseline

Typical Application Instance	Protocol	Movement Speed	Control Precision	Control Period (ms)	Escape Delay (ms)	Downtime per Year	Reliability	Network Delay		
								RTT (ms)	RTT Jitter (ms)	Max. Transmission Delay (ms)
Remote overhead crane control	S7 com	< 1.2 m/s	< 1.0 m	400	1200	< 0.9h	> 99.99%	< 50	< 12.5	< 1200

Typical Application Instance	Protocol	Movement Speed	Control Precision	Control Period (ms)	Escape Delay (ms)	Downtime per Year	Reliability	Network Delay		
								RTT (ms)	RTT Jitter (ms)	Max. Transmission Delay (ms)
Unmanned overhead crane	S7 com	< 0.6m/s	< 0.2 m	80	240	< 0.9h	> 99.99%	< 20	< 5	< 240

- Reliability:

  According to 3GPP TS 22.104, in the industrial scenario, the reliability must be 99.99%, the difference in delay between transmission intervals cannot exceed 25%, and the escape delay cannot be less than triple the transmission interval delay.

  Reliability = MTBF / (MTBF + MTTR) = 0.9 / (365 *24) = 99.99%. Therefore, the downtime per year is less than 0.9 hours.

  When the maximum transmission delay exceeds the PLC escape delay, the service triggers the escape mechanism, shutting down the device and reverting it to its preset safe settings, which interrupts the service.

- Remote overhead crane scenario: Control precision = Device movement speed x RT_e2e

  For example, if the movement speed is 1.2 m/s and the safe distance is 1.0 m, the delay is 830 ms (1.0 m /1.2 m/s = 830 ms).

  It can be calculated that the network RTT requirement is 50 ms: 830 (total delay) – 20 (average sample delay) – 277 (encoding and decoding delay) – 8 (average refresh delay) – 205 (average personnel response delay) – 200 (average PLC message sending delay) – 70 (mechanical delay) = 50.

- Unmanned overhead crane scenario: Control precision = Device movement speed x PLC activation delay

  For example, if the movement speed of an overhead crane is 0.6 m/s, the relative movement speed of two overhead cranes is 1.2 m/s, and the safe distance is 0.2 m, the value of PLC activation delay is approximately 0.17s (0.2/1.2 ≈ 0.17).

  It can be calculated that the network RTT in this case is 20 ms: 170 (total delay) – 70 (mechanical transmission delay) – 80 (maximum PLC transmission delay) = 20.

7.5 Massive Connectivity Service

7.5.1 Impact Factor

The mMTC service usually requires a small data packet size and little throughput, but the density of devices is high. The biggest challenge is the network's capability to support high connection density and network energy efficiency. In most cases, the quality of the service is evaluated in terms of service experience and transmission quality. However, the mMTC service has low requirements on transmission, but high requirements on network resource consumption and network energy consumption utilization. From the perspective of network construction, the evaluation objective is to ensure good service quality, and maintain minimum resource usage and energy consumption.

Table 7-19 Core factors affecting massive connectivity services

Factor	Impact
Connection density	The connection density refers to the total number of devices that reach the target QoS per unit area (per square kilometer). Reaching the target QoS ensures that the system's packet loss rate is less than 1% under the given packet arrival rate *L* and packet size *S*. Packet loss rate = (Number of interrupted packets)/(Number of generated packets). If the target packet is lost after the packet discarding timer expires, the packet enters the interruption state.
Network energy efficiency	The capability minimizes the energy consumption of the RAN while providing a much better area traffic capacity. Because the density of devices is high, the requirements on energy efficiency are also high. The network energy efficiency of access devices for the service needs to be comprehensively evaluated to appropriately distribute devices and ensure that their energy efficiency levels meet requirements for network construction. In addition, the relationship between the device's behavior and network energy efficiency can be analyzed to appropriately plan the service behavior of devices and maximize network energy efficiency.
Packet loss rate	High-density mMTC services do not have high requirements on bandwidth and delay, but require that data packets be successfully transmitted. Therefore, the packet loss rate determines whether the QoS of mMTC services meets requirements.
Accessibility	Network accessibility is key to the performance of the mMTC service. The accessibility refers to the access capability of wireless and core networks. The networks require only normal data transmission as long as access is successful.

7.5.2 Indicator System

Table 7-20 Service indicator system of massive connectivity services

Category	Indicator
Comprehensive service quality evaluation	**Q-score (mMTC)**
Density	Connection Density
	Concurrent Activation Services
Energy efficiency	Network Energy Efficiency Index
Quality	Packet loss rate
	Access success rate

7.5.3 Modeling Method

The Q-score (mMTC) is evaluated from the perspective of quality rather than density or energy efficiency.

The Q-score modeling method is as follows:

Accessibility is a basic factor that affects service availability, and the packet loss rate affects performance.

Assuming that the access success rate is ASR and the packet loss rate is PLR:

$$QScore = ASR \cdot f(PLR)$$

According to Weber's Law, the relationship between PLR and Q-Score complies with the IQX assumption. After f is expanded, the formula is as follows:

$$QScore = ASR \cdot e^{-\alpha \cdot (PLC+\beta)}$$

7.5.3.1 Connection Density

Each device in an activated state interacts with the network side. Therefore, the activation state of the device can be evaluated on the network side, and the number of activated devices and initiated connections can be identified.

The service coverage scope can be obtained as follows:

1. Obtain the baseline data from the wireless engineering parameters to estimate the inter-site distance and further estimate the coverage scope of each baseline and the total coverage scope. In this way, the number of connections can be associated with each base station and area.
2. Obtain the coverage scope from MRs. In most cases, MRs do not carry information on the longitude and latitude, and the penetration rate is low, which is the reason that this method is not often used.
3. Obtain information on the longitude and latitude from the data reported by devices, which is dependent on the design of the service. There is no GPS signal and longitude or latitude information is not reported in the indoor scenario. In addition, taking into account the security and performance overhead, longitude or latitude information may not be reported. Therefore, this method is not widely used.

NOTE

The massive connectivity service has high requirements on the energy efficiency of terminals. Therefore, it is unlikely to embed quality measurement software in terminals.

7.5.3.2 Concurrent Activation Services

Connection Density is used to measure the capability of a 5G network to support the massive connectivity service. However, terminals accessing the network may not initiate services concurrently. A large number of terminals simultaneously initiating evaluation requests to the network places huge demands on the network, and massive user connections are generated instantaneously, which may cause access failures and network congestion.

Concurrent Activation Services indicates the total number of service lines that can be detected in a statistical period.

This KPI reflects the concurrent performance requirements of the massive connectivity service on the network and helps detect unexpected fluctuations in traffic.

7.5.3.3 Network Energy Efficiency Index

[Background]

At the 3GPP RAN # 72 meeting held in June 2016, two network performance indicators were added. One is **Network Energy Efficiency** (for details, see 3GPP TR38.913), which became a network performance indicator of 5G networks. For details, see section 2.2.

According to 3GPP TR 38.913:

$$EE_{global} = \sum_{scenario\ K} b_K EE_{scenario\ K}$$

Among which b_k indicates the weight of each deployment scenario for evaluating network energy efficiency. The unit is FFS.

$$EE_{Scenario} = \sum_{load\ level\ 1} a_1 \frac{V_1}{EC_1}$$

- V_1 indicates the traffic per second (unit: bps) of base station services.
- EC_1 indicates the power consumed by a base station to provide V_1 services (unit: watt = joule/s)
- a_1 indicates the weight of each traffic load level. The unit is FFS.

For the ToB network energy efficiency index, both the baseline energy consumption and the impact of terminals on network energy consumption must be taken into account. For the IoT, the number of things is not limited, and their energy consumption requirements are very stringent. Various factors are considered by enterprise tenants in IoT scenarios, including: whether the network consumption differs from B2C scenarios after a large number of things are connected; the impact on the network; whether the power consumption is appropriate when services are provided on the 5G network; and the price-performance ratio. Network operators need to take into account whether the impact of massive access on the network is controllable. On the premise of meeting the QoS requirements of terminals, they not only need to maximize spectral efficiency, but also need to consider the balance between spectral and network efficiency.

NOTE

The information of power consumption can be obtained from the statistics of base station traffic.

7.5.4 Experience Baseline

Dependent on project requirements.

7.6 FWA Service

7.6.1 Impact Factor

Table 7-21 Core factors affecting IP voice and videoconferencing services

Factor	Impact
Packet loss rate	Packet loss can cause issues in voice quality, such as unclear speech and discontinuity.
One-way delay	One-way delay refers to the delay from voice packet sending to receiving. Long delay causes the delay in speech, affecting user experience.

Factor	Impact
Jitter	Jitter refers to the change of delay on the network side. A small jitter can be eliminated by the jitter buffer on terminals. However, the elimination of jitter causes delay, and large jitters cannot be eliminated, which causes the voice quality to fluctuate and affects call experience.

7.6.2 Indicator System

The enterprise FWA service indicators are the same as those of traditional services. You are advised to obtain the indicators from the network management system of the core network or AR routers. In terms of pipe transmission capability, the SLA of the IP network needs to be guaranteed based on the type of service.

The following figure shows FWA networking.

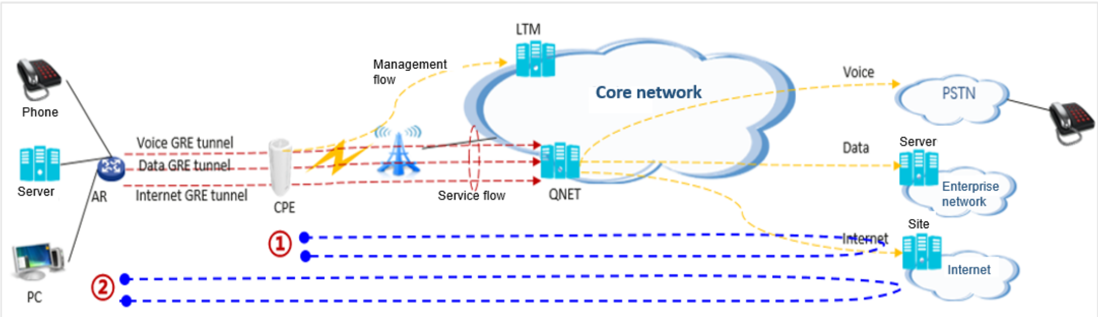


Table 7-22 FWA service indicator system

Indicator	Object Measured	Measurement Method	SLA Threshold
Availability	CPE	SoC monitors the user plane. If an IP flow does not transmit data within a specific period of time, SoC invokes CPE IP Ping through the LTM interface to ping the destination IP address. If the test fails, it indicates that a fault has occurred.	99.995% (dependent on enterprise requirements)
Delay and packet loss rate	CPE/AR	AR launches a ping test at regular intervals. SoC invokes CPE IP Ping through the LTM interface to ping the destination IP address.	0.5% (dependent on enterprise requirements) 25 ms (dependent on enterprise requirements)

Indicator	Object Measured	Measurement Method	SLA Threshold
Rate (uplink rate, downlink rate)	CPE	SoC invokes CPE IP Ping (TR.143) through the LTM interface to measure the uplink and downlink speeds of the destination IP address.	10M UL, 20M DL (dependent on enterprise requirements)

Enterprise FWA service experience modeling: Analyze mainstream services on enterprise networks, including IP voice services, IP videoconferencing services, and Internet services (such as email sending and receiving, VoD, browsing, and VoIP). Enterprise data services, such as fax services, have low requirements on network bandwidth and delay. Therefore, no special modeling or assurance is required.

Table 7-23 FWA service indicator system

Service Category	Indicator
IP voice and videoconferencing	Uplink throughput
	Downlink throughput
	Packet loss rate
	One-way delay
	Jitter
Internet – email service	Uplink throughput
	Downlink throughput
Internet – web browsing service	Downlink throughput
Internet – video service	Xkb start delay
	Downlink throughput
	Downlink RTT
	Throughput bitrate ratio
Internet – VoIP service	Uplink throughput
	Downlink throughput
	Packet loss rate
	One-way delay
	Jitter

- Characteristics of voice services:
  - SIP messages on the signaling plane and UDP/RTP messages on the user plane are transmitted in small packets bidirectionally.
  - Voice services have low requirements on bandwidth. The key factors that affect service experience are packet loss rate, delay, and jitter.
  - The signaling-plane call completion rate, call connection delay, and call drop rate are related to enterprise gateways but not FWA service pipes. Therefore, you are not advised to monitor these indicators on the CEM platform.
- Characteristics of video conferencing services:
  - SIP messages on the signaling plane and UDP/RTP messages on the user plane are transmitted in large packets bidirectionally.
  - Sufficient bandwidth must be ensured. The key factors that affect video conferencing service experience are packet loss rate, delay, and jitter.
  - The signaling-plane call completion rate, call connection delay, and call drop rate are not closely related to FWA service pipes. Therefore, you are not advised to monitor these indicators on the CEM platform.

7.6.3 Modeling Method

FWA is a networking method. Therefore, the existing service modeling systems can be reused. For example, the existing SIP voice service indicator system can be used for FWA SIP calls, and the existing web page, video, email, and VoIP service indicator system can be used for VoIP services.

7.6.4 Experience Baseline

[IP voice service]

The common codec formats for enterprise networks are G.729 and G.711, and the effective rates are 34.4 kbps and 90.4 kbps.

Codec Format	Throughput (kbps)	Codec Format	Rate (kbps)
G.729	34.4	G.723.1	20
G.711	90.4	iLBC	28

The following table lists the requirements for voice service network indicators.

Grade	Delay (ms)	Jitter (ms)	Packet Loss Rate (%)
Good	≤ 40	≤ 10	≤ 0.2%
Fair	≤ 100	≤ 20	≤ 1%
Poor	≤ 400	≤ 60	≤ 5%

NOTE

The preceding requirements are determined based on *YD/T1071-2000 Technical Requirements for IP Telephony Gateway* and China Mobile's VoLTE assessment standards. The packet loss rate in the preceding table is the two-way packet loss rate.

[IP video conferencing service]

The network bandwidth required by video conferencing services is the line bandwidth, which is 1.2 to 1.5 times the conference bandwidth.

Video Format	Recommended Conference Bandwidth		
	H.265	H.264 HP	H.264 BP
720p25/30	768 kbps	768 kbps	1.5 Mbps
720p50/60	1152 kbps	1.5 Mbps	2 Mbps
1080p25/30	1152 kbps	1.5 Mbps	3 Mbps
1080p50/60	2 Mbps	3 Mbps	4 Mbps
4K30	5 Mbps	–	–

Grading Criteria	Delay (ms)	Jitter (ms)	Packet Loss Rate (%)
Good. The video is smooth and clear; the audio is clear.	≤ 100	≤ 30	≤ 1%
Fair. The video is smooth most of the time, but slight pixelation occurs in large movement scenarios.	≤ 100	≤ 30	≤ 3%
Poor. The video is smooth most of the time, but slight pixelation and frame freezing occur in large movement scenarios, which may cause delay.	≤ 100	≤ 50	≤ 5%

[Internet – email services]

Baseline reference:

Uplink Rate	Downlink Rate
1 Mbps	2 Mbps

For details, visit https://gobrolly.com/data-bandwidth-email-requirements/.

[Internet – web browsing service]

Grade	Loading Delay (s)	Throughput (kbps)
Excellent	< 1	> 8000
Good	< 2	> 4000
Fair	< 3	> 2700
Poor	> 10	< 800

NOTE
The throughput is related to the web page size. The preceding baseline uses a large page (1 MByte) as a reference.

Baseline reference:

https://ieeexplore.ieee.org/abstract/document/6263888/references#references

https://www.hobo-web.co.uk/your-website-design-should-load-in-4-seconds/

[Internet – video service]

Resolution	Throughput (Mbps)	RTT (ms)
480P	0.9	80
720P	2	60
1080P	3.9	45
2K	7.8	30
4K	17.6	20

[Internet – VoIP service]

The common codec formats are SILK (Skype) and Opus (WhatsApp and Facebook_Messenger).

Service Type	Throughput (kbps)
Audio	64
Video	384

Delay	Jitter	Packet Loss Rate
≤ 150	≤ 30	≤ 3%

8 References

1. Video Services Forum, Technical Recommendation TR-06-1:Reliable Internet Streaming Transport (RIST) –Simple rofile,http://vsf.tv/download/technical_recommendations/VSF_TR-06-1_2018_10_17.pdf, 2019-01-18
2. C. Keighrey, R. Flynn, S. Murray, N. Murray, "A qoe evaluation of immersive augmented and virtual reality speech language assessment applications", 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6, May 2017.
3. Recommendation itu-t p.910: Subjective video quality assessment methods for multimedia applications, pp. 1-42, Apr. 2008.
4. B. Bauman, P. Seeling, "Towards still image experience predictions in augmented vision settings", Proc. of the IEEE Consumer Communications and Networking Conference (CCNC), pp. 1-6, Jan. 2017.
5. M. Fiedler, T. Hossfeld, P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service", IEEE Network, vol. 24, no. 2, pp. 36-41, March/April 2010.
6. ETSI TR 126 944 V15.0.0 End-to-end multimedia services performance metrics, 2018-07.
7. Al-Zubaidy H, Liebeherr J, Burchard A. Network-layer performance analysis of multihop fading channels[J]. IEEE/ACM Transactions on Networking (ToN), 2016, 24(1): 204-217.
8. Jiang Y. A basic stochastic network calculus[J]. ACM SIGCOMM Computer Communication Review, 2006, 36(4): 123-134.
9. Jiang Y, Liu Y. Stochastic network calculus[M]. London: Springer, 2008.
10. Xiao C, Zeng J, Ni W, et al. Delay guarantee and effective capacity of downlink NOMA fading channels[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(3): 508-523.
11. Durisi G, Koch T, Popovski P. Toward massive, ultrareliable, and low-latency wireless communication with short packets[J]. Proceedings of the IEEE, 2016, 104(9): 1711-1726.
12. Muhammad Fahim, Alberto Sillitti, Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review. IEEE Access, 10 June 2019, 18770310
13. Profinet industrial communication protocol analysis report
14. 5G Mobile Network Architecture for diverse services, use cases, and applications in 5G and beyond, 5G PPP
15. 5G-ACIA_White_Paper_5G_for_Non-Public_Networks_for_Industrial_Scenarios, 5G ACIA(Alliance for Connected Industries and Automation)
16. NGMN_5G_White_Paper_V1_0, NGMN Alliance.
17. WP_5G_for_Automation_in_Industry, Primary use cases, functions and service requirements.

Abbreviations and Acronyms

Abbreviation	Full Spelling
3GPP	3rd Generation Partnership Project
5G PPP	5G Infrastructure Public Private Partnership
5G-MoNArch	5G Mobile Network Architecture
AGV	automated guided vehicle
AI	artificial intelligence
AIV	air interface variant
ToB	business-to-business
B2C	business-to-consumer
B2H	business-to-home
CDF	cumulative distribution function
CPE	customer-premises equipment
DSCP	differentiated services code point
E2E	end-to-end
eMBB	enhanced Mobile Broadband
FoV	field of view
FPS	frames per second
FWA	fixed wireless access
GOP	group of pictures
HD	high-definition
IMT-2000	International Mobile Telecommunications 2000
IPsec	Internet Protocol Security
iTBR	I frame throughput-to-bitrate ratio
ITU-R	International Telecommunication Union - Radio communication Sector
KPI	key performance indicator
MBB	mobile broadband
MCL	maximum coupling loss
MDT	mean down time
MEC	mobile edge computing
MIT	mobility interruption time

Abbreviation	Full Spelling
mMTC	Massive Machine-Type Communications
MTBF	mean time between failures
MUT	mean up time
NF	network function
OTT	one-trip time
PLC	programmable logic controller
QoE	quality of experience
QoS	quality of service
RAN	radio access network
RIST	Reliable Internet Streaming Transport
RIT	radio interface technology
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
RTT	round-trip time
SDU	service data unit
SLA	Service Level Agreement
SRIT	set of RITs
SRT	Secure Reliable Transport
TRxP	transmission and reception point
UAV	unmanned aerial vehicle
UDT	UDP-based Data Transfer
URLLC	ultra-reliable low-latency communication
V2X	vehicle-to-everything
VM	virtual machine
VNF	virtual network function