iLab

iLab

Huawei iLab

2015 Technical Report

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base Bantian, Longgang Shenzhen 518129 People's Republic of China

Website: http://www.huawei.com

Contents

Huawei iLab is a powerhouse in new fixed-network services and experience research that focuses on infrastructure and the experience of future hotspot services. The research covers trends, algorithms, standards, and tool development, providing the theory to support boosting fixed network industry development and expanding fixed network space.

Huawei iLab will develop architecture research for Network 2020, quantitative analysis on key indicators, and SDN/NFV-based end-to-end ICT integration and third-party authentication, with an aim promote SDN/NFV ecosystem development.

1. 2015 Big Data Analytics Report on China's Internet Videos

In early 2015, the huge potential of China's Internet video industry was demonstrated when Chai Jing's documentary film "Under the Dome" attracted 100 million views in a day. Huawei iLab analyzes the videos provided by China's leading OTTs as well as the videos owned by Sichuan Telecom. Let's take a look at the mission-critical big data behind the analysis.

Key findings

[User behavior] **Mobile terminal + fixed network pipe** are the main channels for OTT video playback. **57%** of VOD requests are from mobile terminals, and **98%** VOD requests are carried by fixed networks. Wi-Fi is still the preferred access mode.

[User behavior] Tier-1 cities including **Beijing, Shanghai, Guangzhou, and Shenzhen** make up approximately **58%** of video views. Video popularity peaks **within one week** with the most views occurring at **22:00.**

[Video content] Internet videos have entered the UHD era. In China, **76%** of videos support **UHD (720p) or a higher** resolution, but commercialized 4K content is limited.

[Video experience] While user experience of watching videos is at the start-up phase, the nation-wide U-vMOS scores in China stay at **2.** The main cause is the **poor video quality** due to low bit rate.

[Key factors] UHD (720p) videos have an average bit rate of less than **1 Mbps** (70% of YouTube's average bit rate, with the standard for best experience as 12 Mbps). Over the last four years, growth of the bit rate has been almost stagnant. Low bit rate reduces requirements on network bandwidth (the current 10 Mbps bandwidth can support smooth playback). Given that the current network bandwidth capability is not a bottleneck, **low bit rate** has been a key factor for enhancing user experience.

[Key factors] RTT is relevant to another key element: initial loading time. **Localized content deployment and reinforced local IDC construction** can effectively reduce the RTT and enhance user experience.

[Content distribution] Content localization can improve user experience. **42%** of the current video resources are within a province, whereas most of video resources are inter-province, which affects user experience.

[Content distribution] With reducing bandwidth costs taken into account, ICPs choose the P2P distribution mode to alleviate server pressures. **38%** of VOD contents are distributed in P2P mode. A high proportion of P2P scheduling affects user experience.

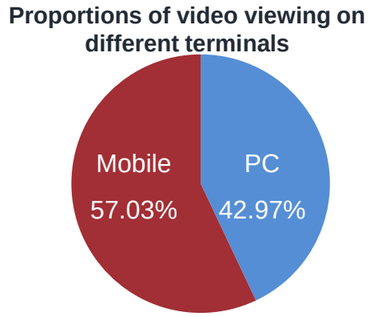
Remarks:

1. The data in this report is obtained using the iLab Spider based on data tracking and in-depth analysis on the open data of videos on China's Internet OTTs including You, Sohu, and iQIYI. A wide range of sample tests are performed among nation-wide video users.
2. The data source must be specified as Huawei iLab.

Most videos are viewed on mobile terminals.

57.03%

57.03% of videos are viewed on mobile terminals, surpassing that on PCs. More and more people are going mobile to watch videos.

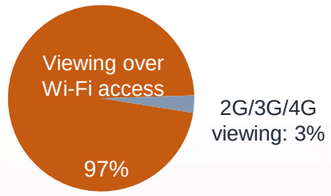


Fixed networks carry over 98% of Internet video traffic.

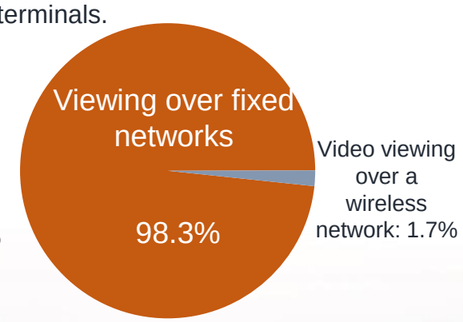
97% of videos viewed on mobile terminals are viewed over Wi-Fi access. Mobile data packages allow users to view videos any time any place over mobile networks, but they are quite expensive.

98.3%

In general, 98.3% of videos are carried over fixed networks. More than 50% of users choose to view videos at home. Public areas with free Wi-Fi are also a preferred choice for users to view Internet videos, especially those who use mobile terminals.



Percentage of video viewing on mobile terminals

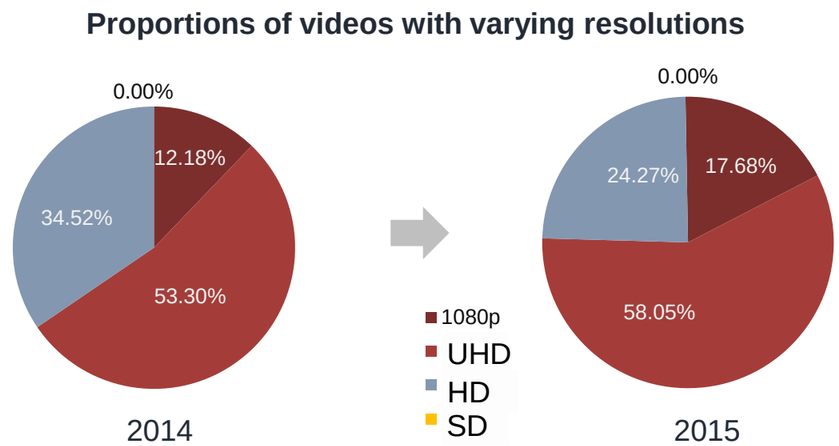


Percentage of video viewing over fixed networks

UHD videos see ever-increasing demands.

76%

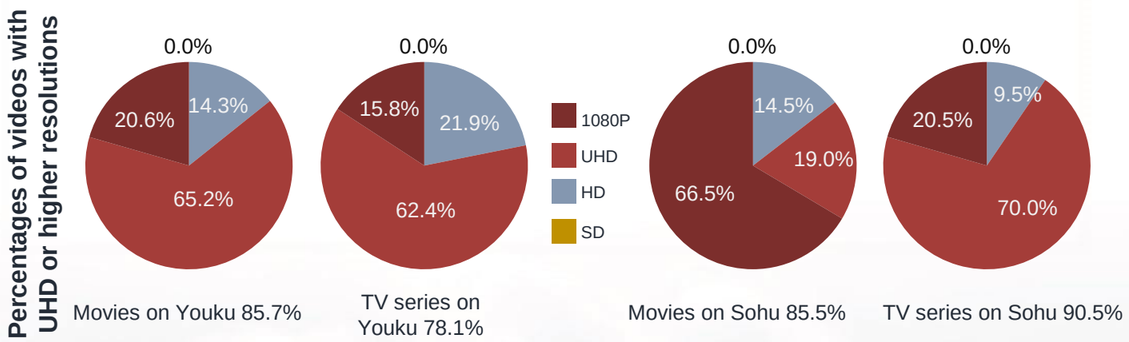
In 2015, 76 out of every 100 newly loaded videos support a resolution of UHD (720p) or higher.



Remarks: The mainstream OTTs in China define the 720p resolution as UHD.

TV series/movie videos with a UHD (720p) resolution or higher account for 90% of the total.

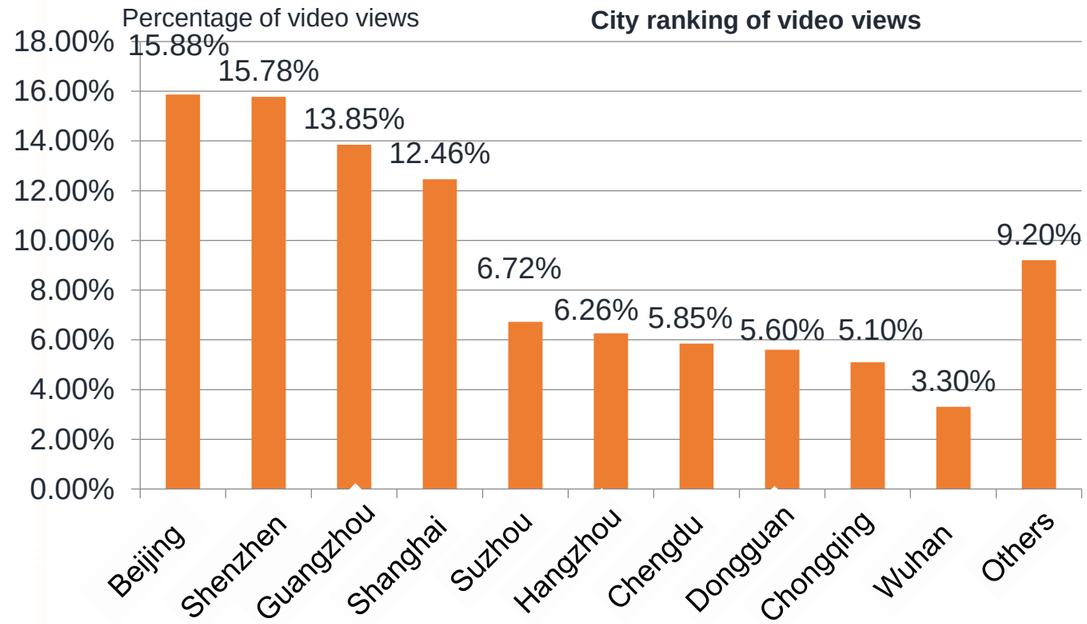
TV series videos and movie videos have entered the **UHD era** and are moving towards the 1080p era.



Tier-1 cities contribute to 58% of video views.

58%

Tier-1 cities (Beijing, Shanghai, Guangdong, and Shenzhen) contribute to 58% of video views. The Internet video user base is mainly located in densely populated and economically developed regions.



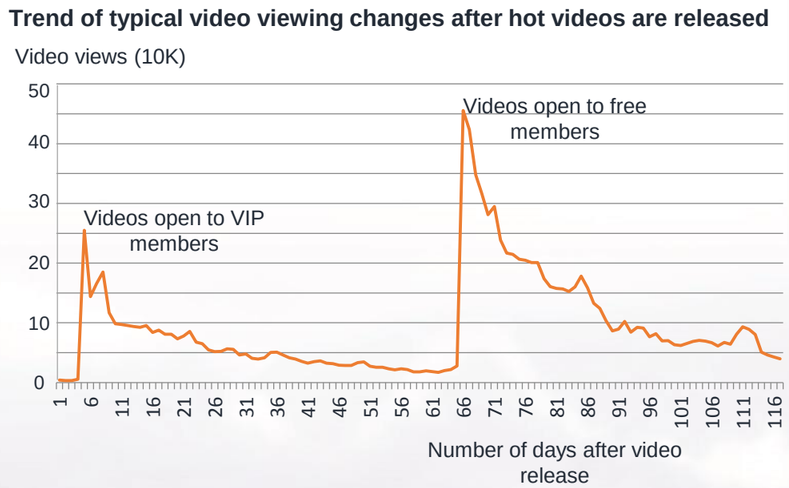
Videos popularity peaks within one week of release

Hot videos encounter two peaks after release.

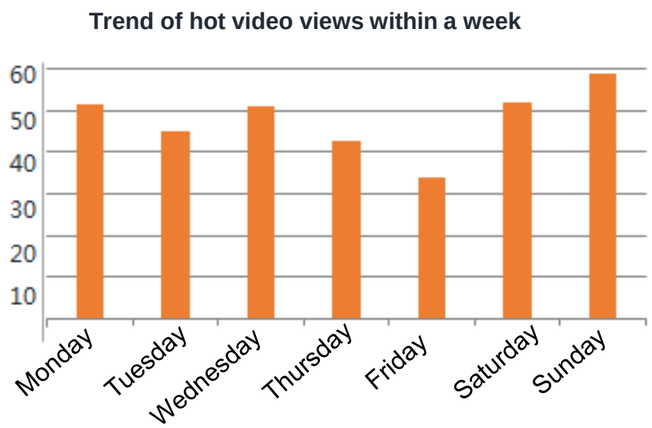
- Within one week after they are opened to VIP members
- Within one week after they are opened to free members

One week after videos are released, the number of video views decreases exponentially. The number of views per day stabilizes after another two months.

The peak value of video views for free members is almost double that for VIP members.

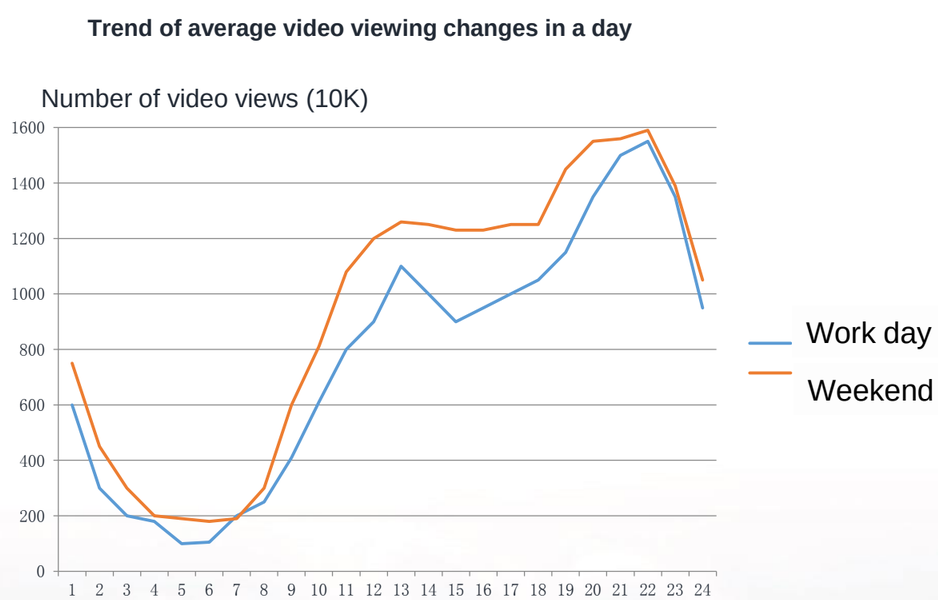


Most videos are viewed during the weekend.



After video viewing enters a grace period, the number of hot video views on Saturday and Sunday peaks within a week. This number then markedly decreases from Monday.

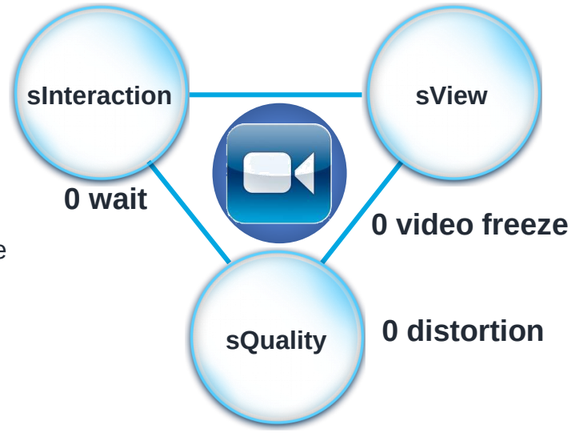
The number of video views peaks at 22:00.



Video views peak at 12:00 and 22:00 each day. Each hour, more videos are viewed during the weekend than during work days.

U-vMOS, Huawei's video experience measurement system

U-vMOS is a video experience measurement system developed by Huawei to boost open industry chain cooperation and continuous evolution. Video experience is scored based on three indicators: video quality, interactive experience, and viewing experience. If any of the three indicators has a low score, the final U-vMOS score will be low, which is a ceiling effect.



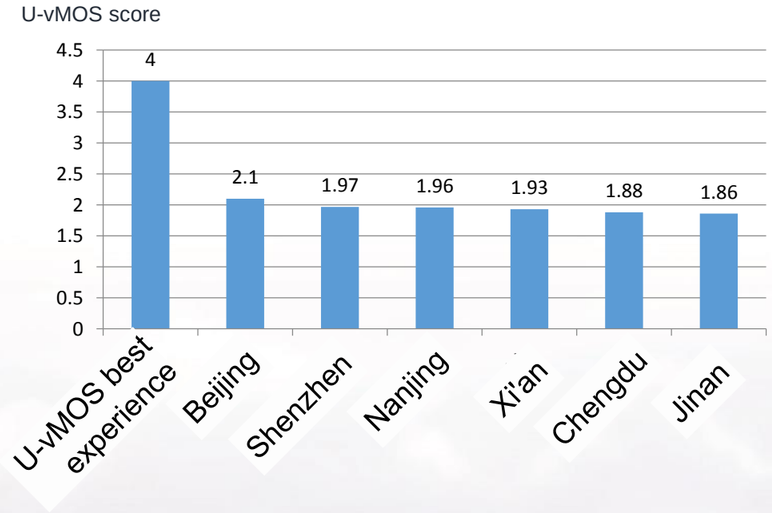
$$U\_vMOS = (sQuality - 1) \cdot \left(\frac{\alpha \cdot (sInteraction - 1) + \beta\ (sView - 1)}{4 \cdot (\alpha + \beta)} \right) + 1$$

Remarks: α and β are the weight values of sInteraction and sView, respectively. The values vary with the service type.

The Internet video experience of sample cities has an average score of 2 (5 as the highest score)

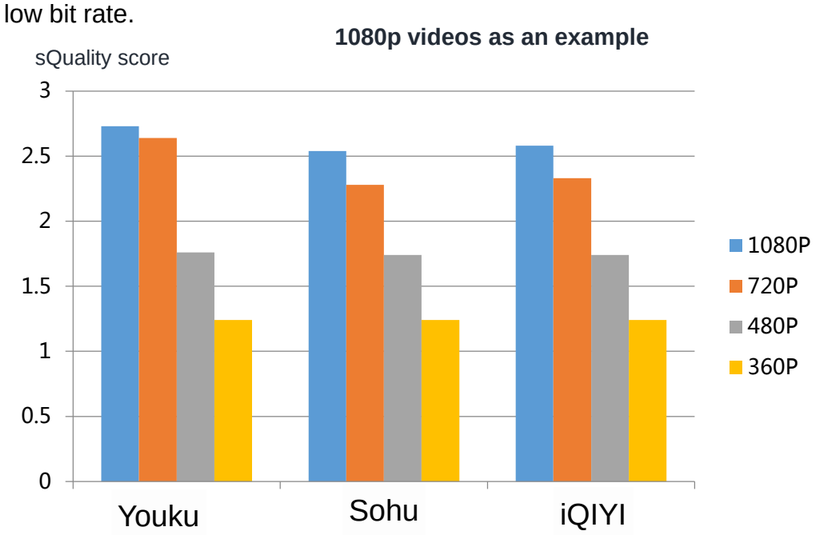
Superior experience: 5 score, initial loading time < 100 ms, 0 video freeze

Best experience: 4 score, initial loading time < 1s, video freeze < 3s per hour



Video quality is the major weakness, with the score under 2.8.

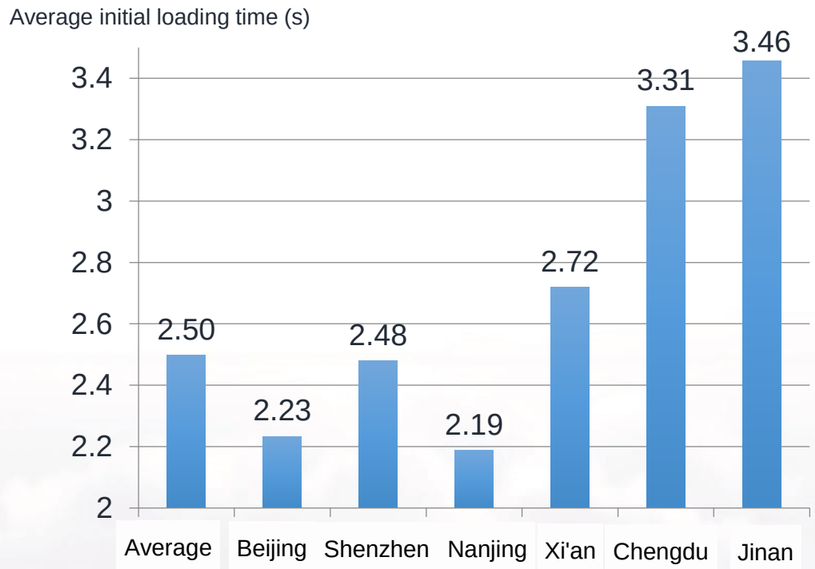
The sQuality scores are unsatisfactory as the ICP program sources of any resolution have a low bit rate.



Remarks: The ICP program source samples tested for all cities are the same; therefore, the sQuality scores for the cities are the same.

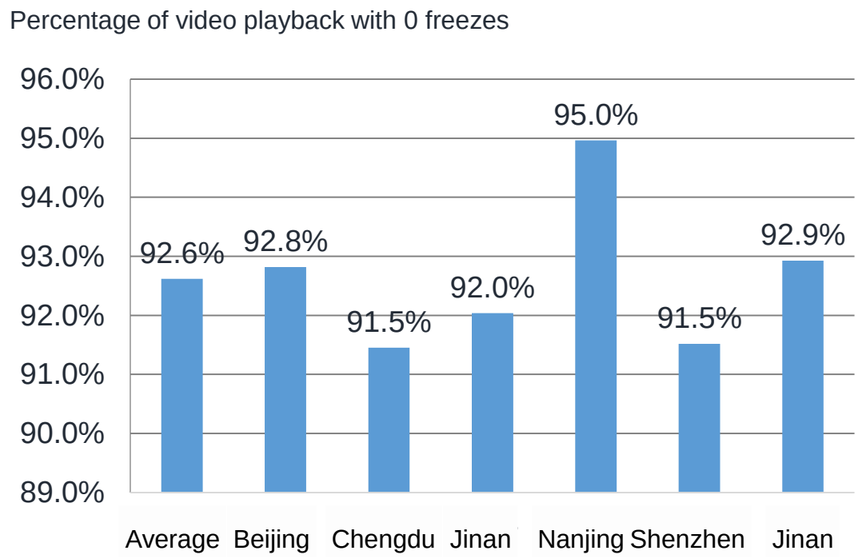
There is a big gap between the initial loading time in various cities and the average value (2.5s).

The initial loading times for video viewing in Beijing, Shenzhen, and Nanjing are lower than the average value. The initial loading times for video viewing in Xi'an, Chengdu, and Jinan are very long.



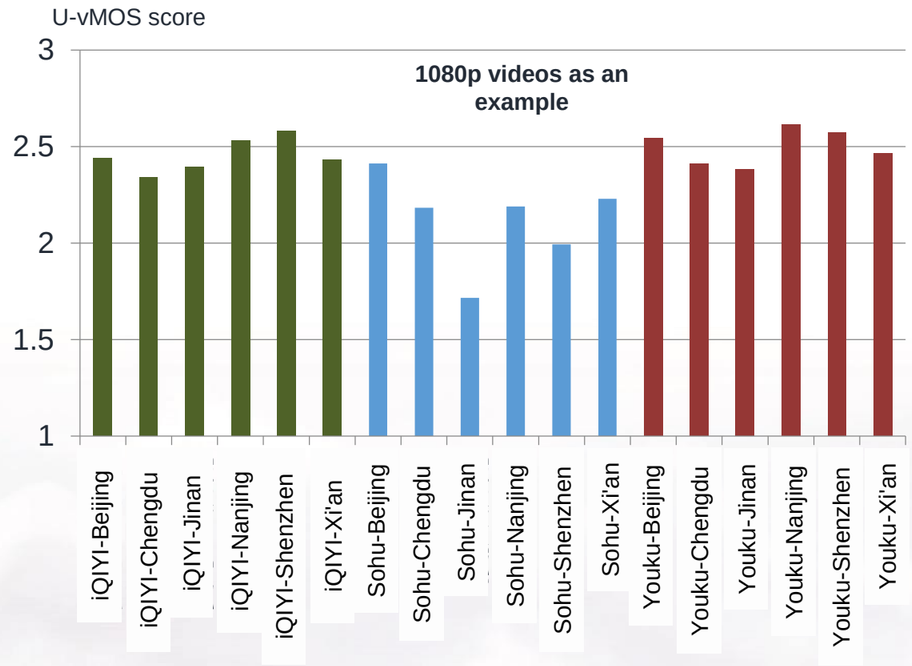
There is a big gap between the initial loading time in various cities and the average value (2.5s).

The initial loading times for video viewing in Beijing, Shenzhen, and Nanjing are lower than the average value. The initial loading times for video viewing in Xi'an, Chengdu, and Jinan are very long.



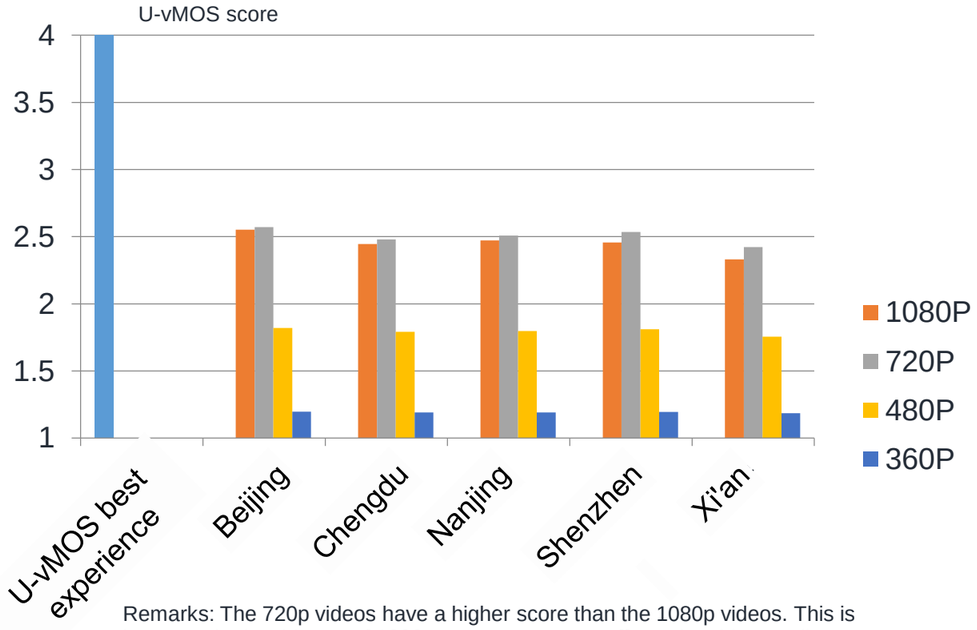
Youku and iQIYI videos have better user experience

U-vMOS scores: Youku 2.52, iQIYI 2.43, Sohu 2.17



The U-vMOS score is in direct proportion to the video quality.

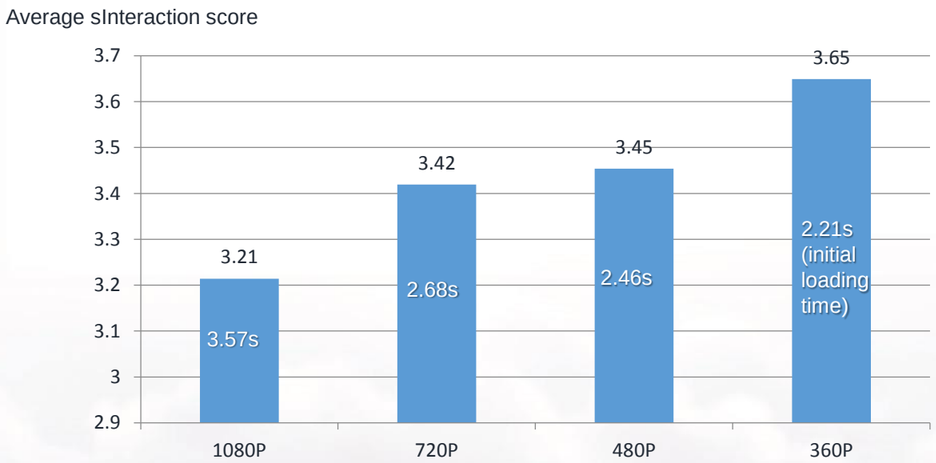
The U-vMOS score is determined by the video quality of program sources. The higher the resolution and bit rate, the better user experience.



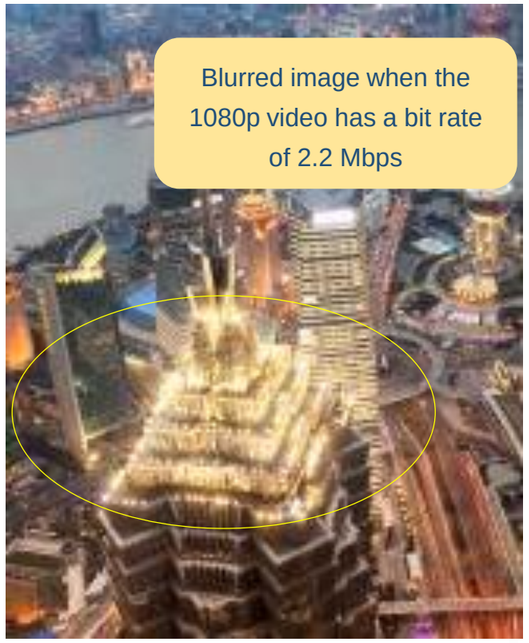
Remarks: The 720p videos have a higher score than the 1080p videos. This is because the low bit rate of 1080p program sources affects the video quality.

The higher resolution of program source, the lower score of the initial loading time (sInteraction).

The higher the resolution of a program source, the longer the initial loading time, and the lower the score.

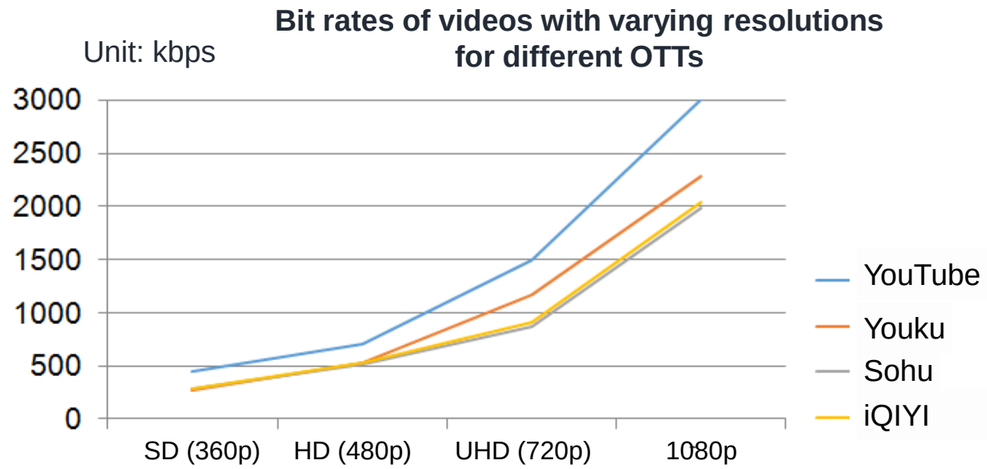


Videos with the same resolution but different bit rates bring varying user experience.



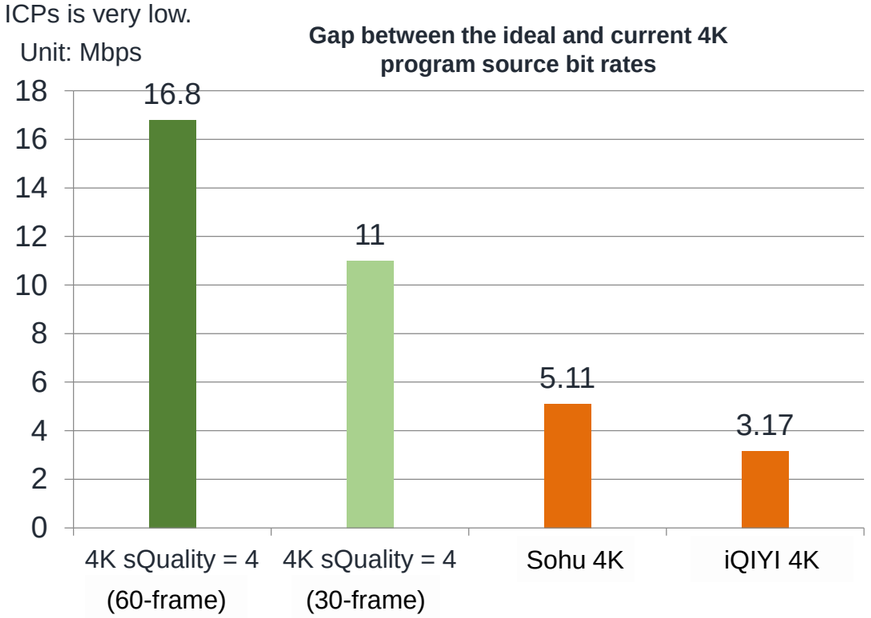
Given the same resolution, Internet OTT videos provided by China's ICPs have a 30% lower average bit rate than videos on YouTube.

Low bit rate is the **same choice** for mainstream OTTs in China.

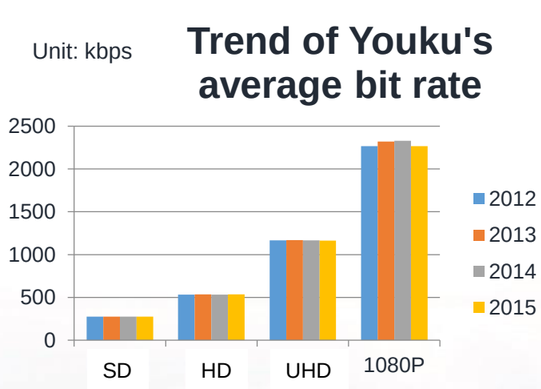


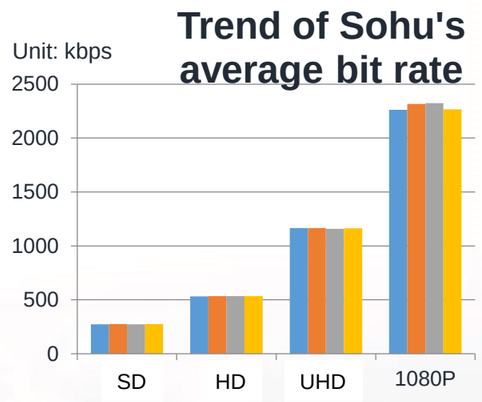
Big gap between the ideal and current bit rates of 4K program sources in China

Among China's ICPs, only Sohu and iQIYI provide a small number of PC-supported 4K program sources. There is a big gap between the ideal and current bit rates.
The human factors engineering test result shows that the typical bit rate of 4K videos with good user experience is 16.8 Mbps. At present, the bit rate of program sources provided by China's ICPs is very low.



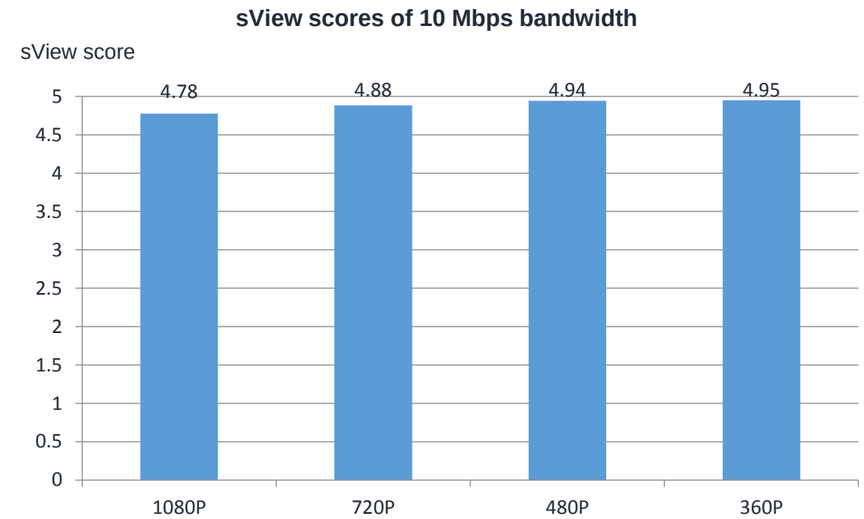
The Internet video bit rate growth stalls.





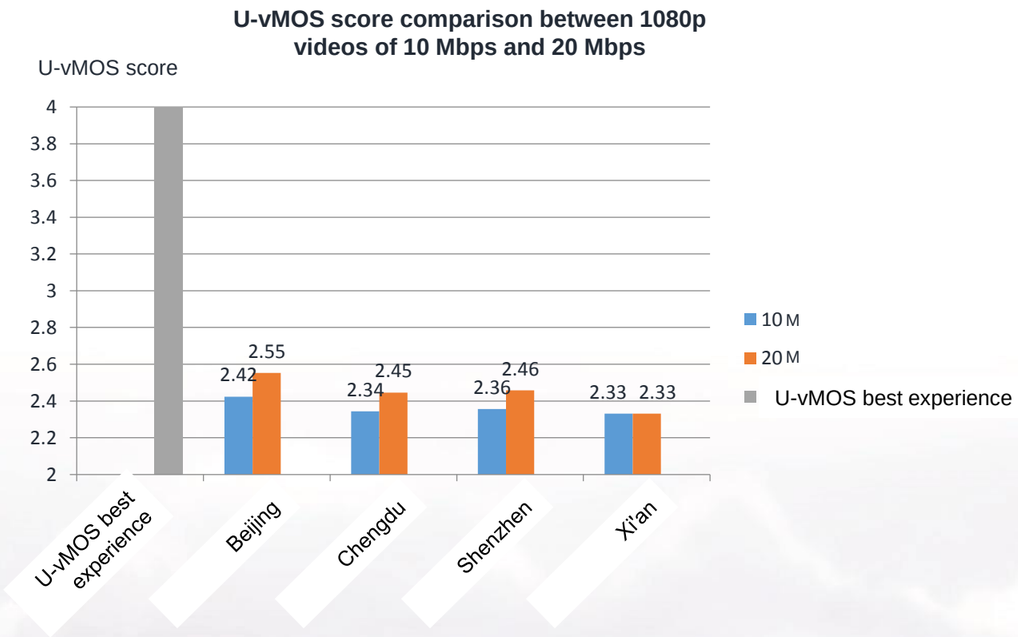
Low bit rate decreases the video quality and requires less network bandwidth but brings higher video fluency.

Currently, most videos have a low bit rate. Generally, videos can be played smoothly at a 10 Mbps bandwidth. Low video quality is an alternative to high video fluency.



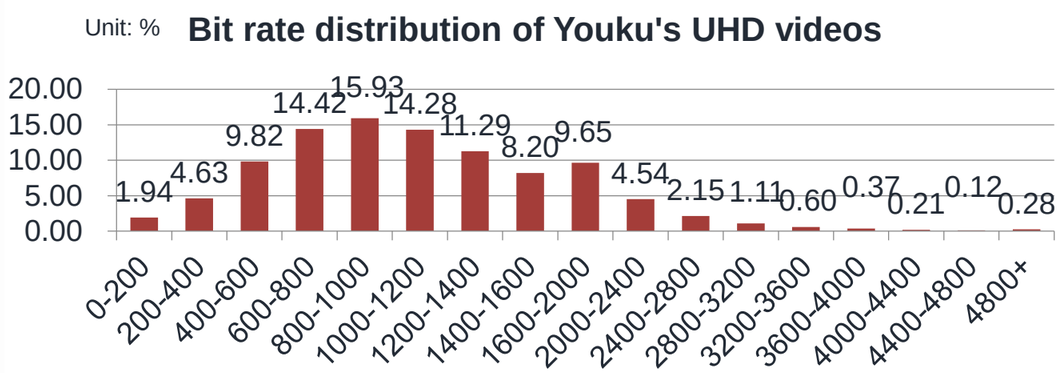
Higher bandwidth can improve user experience. When bandwidth is not a bottleneck, a higher bit rate is required to improve video quality.

With low bit rate videos, increasing the network bandwidth from 10 Mbps to 20 Mbps will have a limited effect on improving user experience.



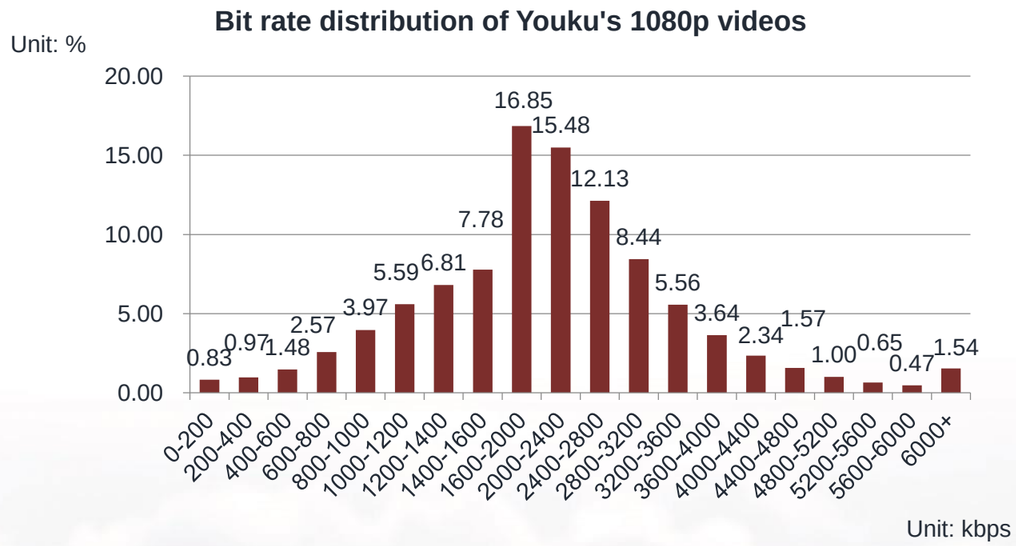
The bit rate of Youku's UHD videos ranges from 400 kbps to 2000 kbps.

400 to 2000 kbps for UHD



The bit rate of Youku's 1080p videos ranges from 1.0 Mbps to 3.6 Mbps.

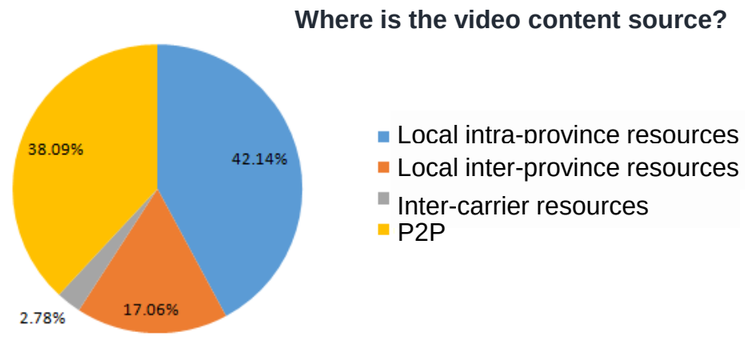
1.0 to 3.6 Mbps for 1080p



Only 42.14% of video resources are from the local or provincial network.

42.14% Only 42.14% of the program source contents are from the IDCs of OTT vendors who cooperate with local or provincial network carriers.

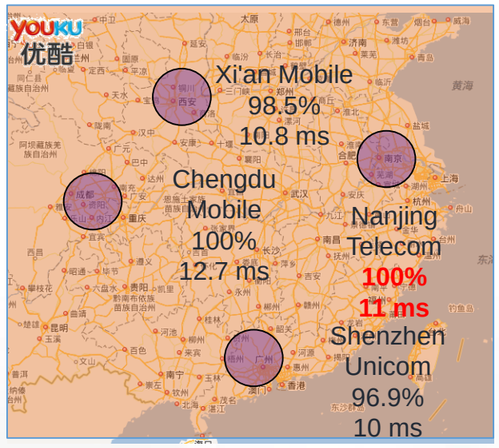
38.09% of video contents are distributed in P2P mode. With reducing bandwidth costs taken into account, ICPs prefer the P2P distribution mode to alleviate the server pressure.

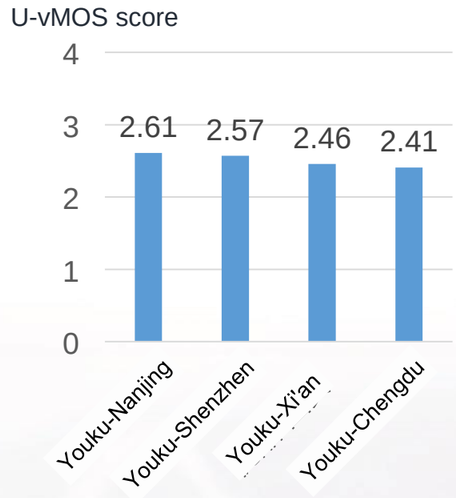


The higher the content localization rate, the less RTT and better initial loading experience.

Higher content localization rate together with excellent IDC quality can improve OTT video experience. In provinces that have a high content localization rate for videos provided by China's top 3 carriers, the average RTT is 10 ms, and the U-vMOS scores are high.

In 2015, Nanjing Telecom had 100% of localized scheduling for Youku's video contents and excellent metro network IDC quality, with the overall video experience ranking number one.





U-vMOS scores for Youku 1080p videos

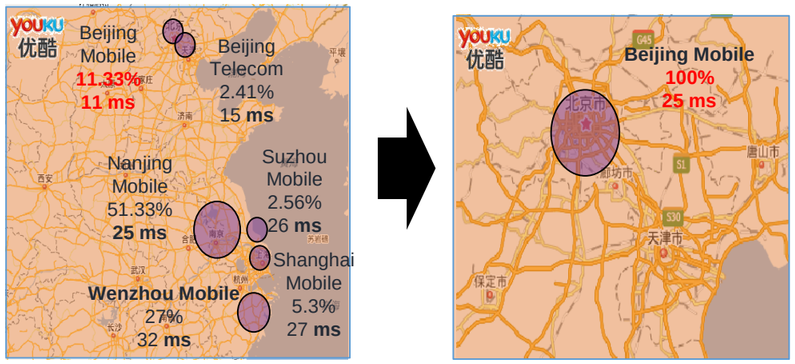
Remarks: The percentage indicates the content localization rate.

Content localization does not mean zero congestion. More effort needs to be made on IDC network construction to improve video experience.

Beijing Mobile is improving content localization, and in 2015, **Youku videos** reached 100% of local scheduling. However, after content localization, the increased local traffic led to IDC network congestion and local network KPI deterioration, resulting in only **limited improvements** to the overall Youku video experience.

2014 content localization rate: 11.33%
Local RTT: 11 ms

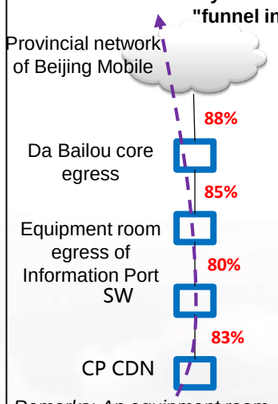
2015 content localization rate: 100%
Local RTT: 25 ms



Youku video experience	Experience KQI						Statistics	
	Initial Buffering Time (s)		Number of Video Freezes		Video Freeze Period (s)		Zero Wait Rate	
	Off-peak hours	Peak hours	Off-peak hours	Peak hours	Off-peak hours	Peak hours	Off-peak hours	Peak hours
2014	2.39	2.97	0.05	0.1	0.34	0.75	87.71%	81.55%
2015	1.98	2.24	0.05	0.05	0.25	0.46	90.10%	85.72%

The six-month sample tests and network-wide system assessment helped Beijing Mobile to discover multiple IDC network egress bottlenecks that led to KPI deterioration of the local network. Beijing Mobile has started to address the problems in order to improve Internet video experience.

The IDC network egress does not match incrementally added content, forming a "funnel interface".



Remarks: An equipment room path is used as an example.

Service experience awareness severely deteriorates during peak hours.

The link congestion is heavy during peak hours on various IDC network nodes.

IDC network architecture construction lags behind cloud DC demands

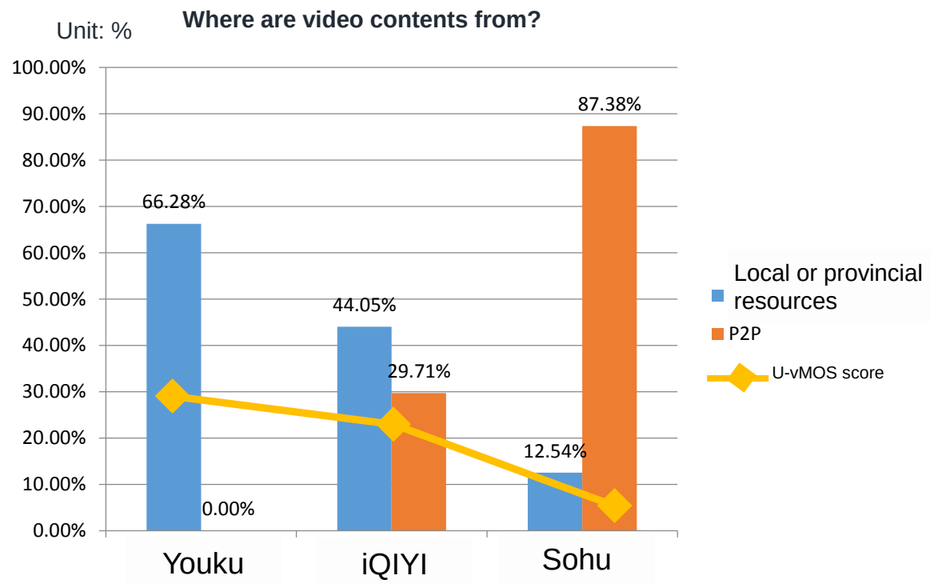
Status Quo	Problems and Impacts
Four-layer architecture (access + aggregation + core + egress)	Multiple layers lead to heavy workload in O&M.
Layer 3 to edge TOR switch	The current architecture does not accommodate cluster services and CP data migration.
Some VIP customers directly connecting to core routers	Customer routes and the DC network are closely coupled, and the number of routes defined by CPs is too large.
Limited space, power, and transmission resources for some equipment rooms	The egress capacity and CP capacity fail to be expanded in time.

Youku has more local or provincial resources scheduled, and Sohu has more P2P resources scheduled.

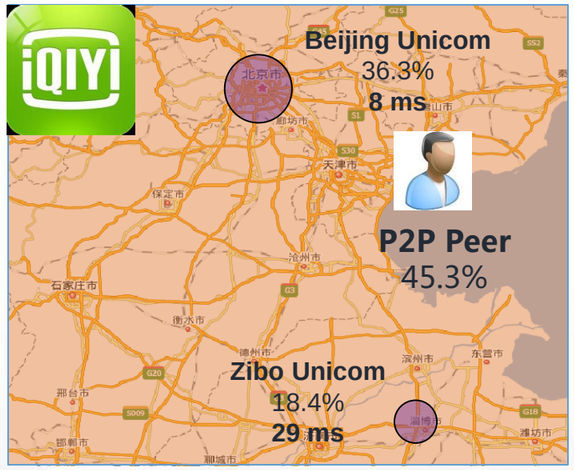
Youku has a higher proportion of local/intra-province resource scheduling.

Sohu has a higher P2P scheduling rate.

U-vMOS scores: 2.52 (Youku), 2.43 (iQIYI), 2.17 (Sohu)



ICPs prefer to use P2P distribution to reduce bandwidth costs.



Video scheduling trend of iQIYI@Beijing Unicom

For them, P2P distribution is used a lot:

Sohu@Beijing Unicom
P2P peer scheduling 96.5%

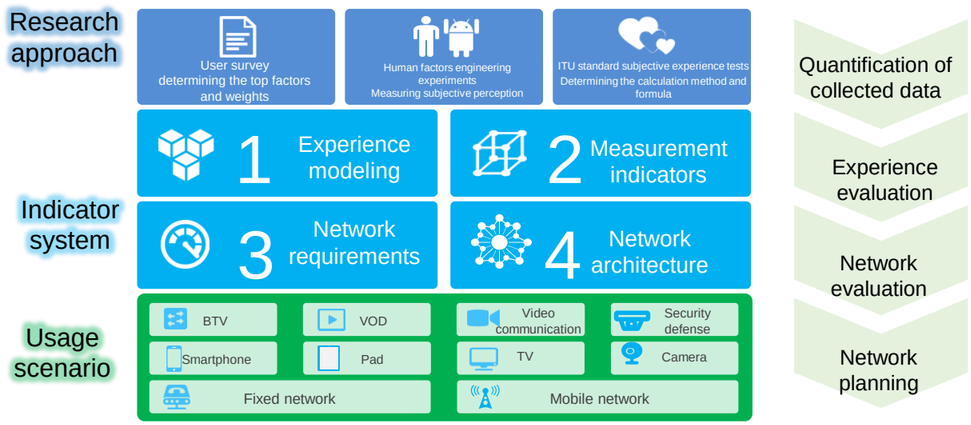
Sohu@Beijing Mobile
P2P peer scheduling rate 34.4%

iQIYI@Chengdu Telecom
P2P peer scheduling rate 34.8%

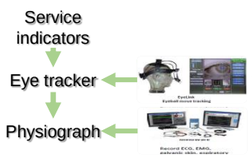
iQIYI@Shenzhen Telecom
P2P peer scheduling rate 50.4%

iQIYI@Xi'an Telecom
P2P peer scheduling rate 42.6%

2. Featured U-vMOS Topics



Human factors engineering experiments (measuring subjective perception)



a. View videos of different qualities on various network conditions.
b. Trace and record the physiological response of test engineers.
c. Obtain the quantitative relationships between perceptions and service indicators based on test engineers' subjective perceptions and perform mathematical modeling.

6000+ samples

ITU standard subjective experience tests (determining the calculation method and formula)

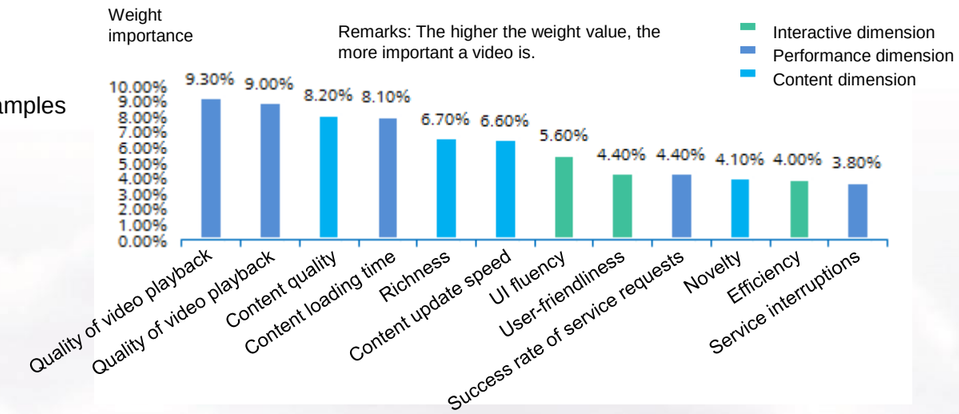
Compliant to
- ITU-R BT. 500
- ITU-T P.910
- ITU-T P.911

6000+ samples

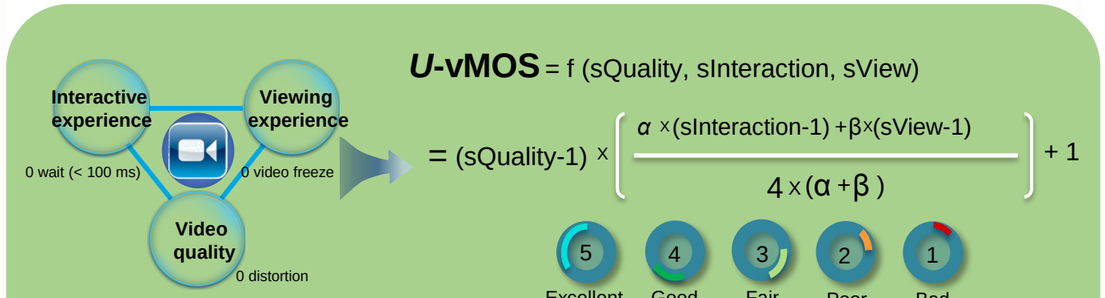
User research (determining top factors and weights)

The top 3 factors influencing user experience are the video source quality, video loading speed, and viewing experience.

3000+ samples



U-vMOS Calculation Formula



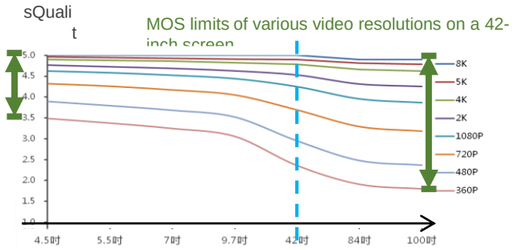
The U-vMOS measures video user experience from the following dimensions:
- Video quality (sQuality): includes video definition, fluency, and fidelity.
- Interactive experience (sInteraction): includes the video loading time and video switching time.
- Viewing experience (sView): includes frame freezing and erratic display.

Remarks: Huawei U-vMOS and its algorithm are jointly produced by Carrier BG, 2012 Laboratory, Fixed Network Product Line, and Wireless Network Product Line.

Video — Quality (sQuality) and interactive experience (sInteraction)

sQuality includes six factors:

- Coding type (such as H.264 and H.265)
- Video bit rate
- Video complexity
- Video resolution
- Video frame rate
- Display size



- The larger the screen, the higher the resolution and the more distinctive experience gaps between resolutions.
- Each resolution has sQuality limits. To improve user experience, increase the resolution.

Maximum sQuality values at different resolutions

Resolution	Screen size						
	4.5-inch	5.5-inch	7-inch	9.7-inch	42-inch	84-inch	100-inch
8K	5.0	5.0	5.0	5.0	5.0	4.9	4.9
5K	4.96	4.95	4.93	4.91	4.90	4.81	4.78
4K	4.90	4.88	4.86	4.82	4.78	4.66	4.62
2K	4.77	4.73	4.69	4.63	4.53	4.31	4.25
1080P	4.62	4.58	4.52	4.44	4.25	3.96	3.87
720P	4.32	4.26	4.17	4.05	3.69	3.29	3.18
480P	3.89	3.79	3.68	3.52	2.95	2.48	2.37
360P	3.49	3.38	3.25	3.06	2.36	1.91	1.80

Standard coverage: 4.5-inch to 100-inch screens

Typical viewing distance: 30 cm (smartphone), 40 cm (PAD), 2.5 m (TV)

sInteraction takes channel change and loading time into account.

BTV scenario

sZapping	
Scale	Channel change delay (ms)
Excellent (5)	<=100
Good (4)	500
Fair (3)	1000
Poor (2)	2000
Bad (1)	>4000

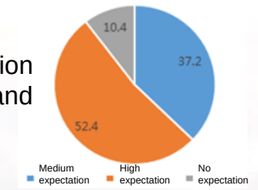
- According to experience, users have high expectations on the channel change delay value.

VOD scenario

sLoading		
Scale	Loading Time (TV)	Loading Time (phone)
Excellent (5)	<=100	<=100
Good (4)	1000	1000
Fair (3)	2000	3000
Poor (2)	5000	5000
Bad (1)	8000	10000

- According to the experience, users have low expectations on the initial loading time.

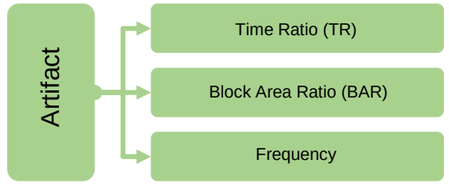
Users have the same expectation on video fluency on both TVs and smartphones.

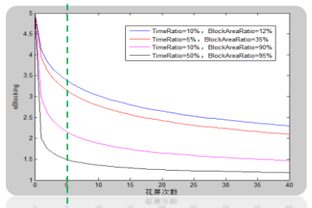


- The loading time for videos on TVs is higher than for videos on smartphones or PADs.
- User expectations on videos vary with the screen. As technologies develop, user expectations on videos on different screens may unify.

Viewing — Experience (sView): BTV artifacts and VOD freezes

BTV scenario:





- Give the same number of BTV artifacts, sBlocking is more relevant to Block Area Ratio and Time Ratio.
- Block Area Ratio has a greater impact on user experience than Time Ratio and Frequency.
- Given the Frequency value ranges from 0 to 5, the user experience is rapidly decreasing.

Typical Values of sBlocking

Score	Time Ratio	Block Area Ratio	Frequency
5	0%	0%	0
4	4%	35%	1
3	10%	45%	2
2	15%	35%	6
1	50%	95%	12

* A 2-hour statistics period is used.

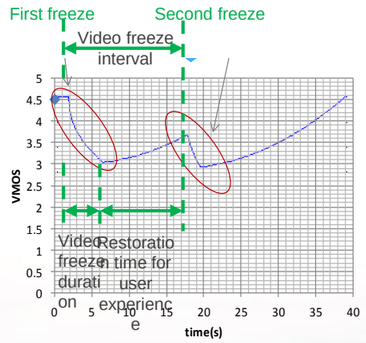
Proportion of Block Area Ratio

VOD scenario:

- The user experience of video freezes is determined by the freeze interval, freeze duration, and freeze frequency.
- The user experience is restored over time between two video freezes.

In a fixed statistical period, Freeze Duration Ratio can be used as a combination of Freeze Duration, Freeze Interval, and Freeze Frequency.

User tolerance for video freezes on TVs is lower than that on smartphones or PADs.



Typical values of video freezes on smartphones or PADs

Score	Number of Times	Average Interval (s)	Average Duration (s)	Time Ratio
5	0	0	0	0%
4	1	0	2.7	5%
3	2	>10	3	10%
2	>2	<5	>5	15%
1	>3	<2	>10	30%

Typical Values of Video Freezes on TVs

Score	Number of Times	Average Interval (s)	Average Duration (s)	Time Ratio
5	0	0	0	0%
4	1	0	2.7	0. 1%
3	3	>30s	9	1%
2	6	>30s	22.5	5%
1	>10	>30s	>27	10%

A one-minute period is used for measuring video freezes on smartphones.

A 45-minute period is used for measuring video freezes on TVs as an episode of a TV drama is typically 45 minutes.

SpeedPro

First fixed network service experience measurement tool

SpeedPro is the first global 4K OTT video U-vMOS measurement tool, supporting one-click smart test on videos, web pages, and bandwidth services.

Video test

U-vMOS
Video freeze times
Download rate

Web page test

Initial display time
Complete loading time
Download rate

Bandwidth test

Delay
Upload rate
Download rate

SpeedPro comes in two versions: PC-professional version and smartphone version.

PC-professional version	VS	Smartphone version
During fixed network evaluation projects, the service experience KQIs, service interaction KPIs, and network KPIs are imported into the uNetBuilder for rapid analysis of the top factors affecting service quality.	Usage scenario	Public users can quickly obtain their quantified service experience quality results on smartphones, and the quality result data is accumulated for user experience analysis.
PC	Running environment	Smartphone
PC requirements: CPU i5, memory 4 GB PC access mode: Network cable access is recommended. If Wi-Fi access is used, the Wi-Fi activation rate must be over 100 Mbps.	Mapping	4K test: Mate8 1080p test: Mate7, P7, Honor 7/7i, Honor 6/6P

FAQ About video experience

Q: Why are the U-vMOS scores not as high as I expected from the video experience?

A: The U-vMOS experience measurement framework is built upon the technical-centric QoE evaluation system, which does not take into account the subjective factors that have individual differences.

In fact, user experience can be described based on the QoE evaluation framework, covering environmental considerations, user considerations, and content considerations. For example, the weather, light, and mood are factored into environmental considerations, and the education level, economic conditions, and the past viewing experience are factored into user considerations.

For example, when it comes to the experience of viewing 1080p videos, a user who had once viewed 4K videos must have different experience compared with a user who

For a description of the tool and FAQs about the tool usage, visit Huawei iLab community at http://3ms.huawei.com/hi/group/2030949.

OTT video service experience research

Buffer is the only perceivable event of OTT video service experience.

A buffer event may be generated during initial video loading or video playback.

Buffer events impact user experience by slowing down video loading or causing video freezes.

Pre-buffering

- Segment download starts when a request for video playback is received.
- When the downloaded data amount reaches a specified buffer threshold, the client starts.
- Pre-buffering is relevant to the initial loading time.

Segment download

- One video is divided into multiple segments that are distributed to different CDN servers for scheduling.
- The video is downloaded and played by segment.
- If video download is not completed within a specified period, video playback will be The download continues when the remaining playing time is less than the pre-buffering threshold.

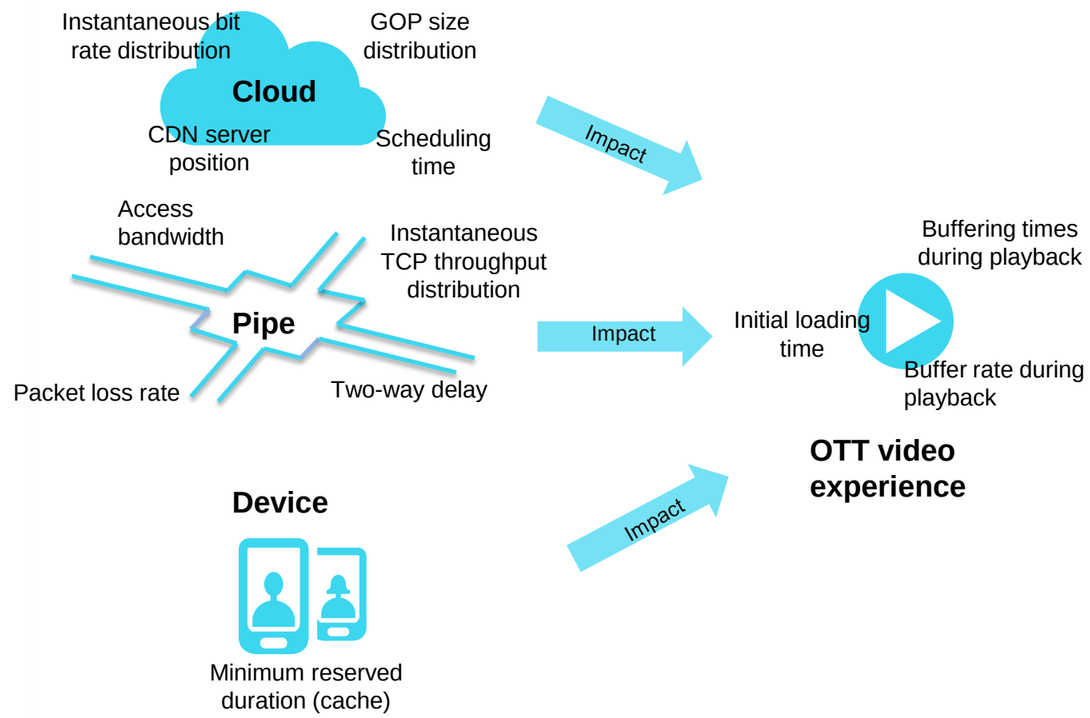
Immediate download

- The client immediately downloads the video while it is playing, making full use of the available bandwidth.
- If the immediate download speed is slower than the current playback progress, a buffer event occurs.

Intermittent download

- The download continues when the remaining playing time is less than the pre-buffering threshold.
- This download can reduce the pressure on CDN server.
- A buffer event occurs if intermittent download is not complete within a specified period.

OTT video experience is determined by various cloud, pipe, and device parameters



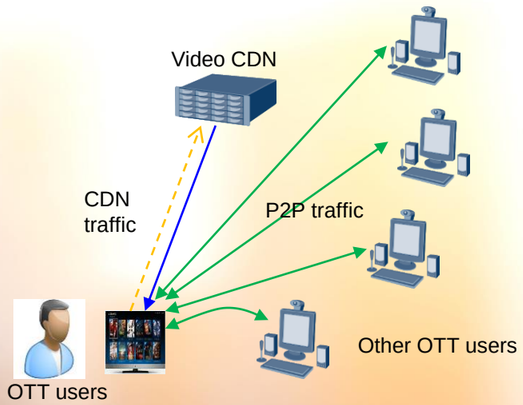
For details, see *Analysis Report on OTT Video Service Experience* released on Huawei iLab community at http://3ms.huawei.com/hi/group/2030949R.

The bit rate is a key characteristic of the OTT video content source and determines the video quality.

- **Official explanation:** Bit rate is the number of bits that are conveyed or processed per unit of time, typically expressed in kbps.
- **Common understanding:** Bit rate is also known as the sampling rate. A greater sampling rate within a specified period indicates a higher precision and that the quality of processed files is closer to that of the original files. Taking an audio file as an example, a higher bit rate means a lower compression ratio, less sound loss, and closer similarity to the source sound.
- However, the file volume is **in direct proportion to** the bit rate. Almost all the coding formats focus on how to achieve the minimal distortion with the lowest bit rate. Such a focus is where CBR (constant bit rate) and VBR (variable bit rate) come from.
- The bit rate determines the quality of OTT video content sources.

The CDN+P2P distribution mode and traffic mode reduce ICPs' storage and bandwidth costs.

For reasons of reducing storage and bandwidth costs, Chinese video ICPs **typically use the CDN+P2P streaming distribution mode**. Given that the user experience is not affected, this distribution mode is preferred for its advantages of saving CDN and bandwidth leasing fees. When the pre-buffering data amount reaches a specified threshold, the client proactively releases CDN connections and changes to use the P2P mode to alleviate the pressure on CDN servers.

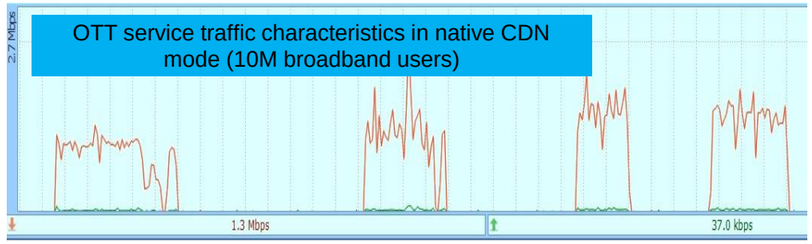


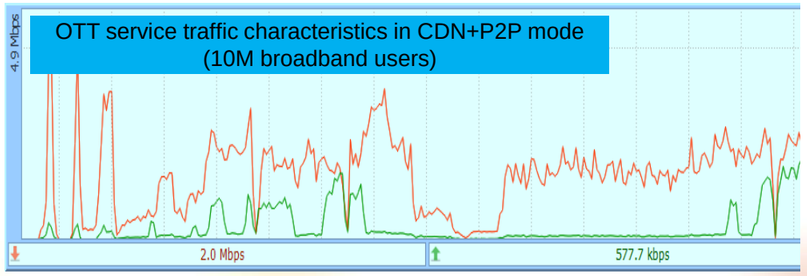
For details, see *Analysis Report on OTT Video Service Experience* released on Huawei iLab community at http://3ms.huawei.com/hi/group/2030949R.

Two streaming distribution modes change the traffic direction of carrier metro networks.

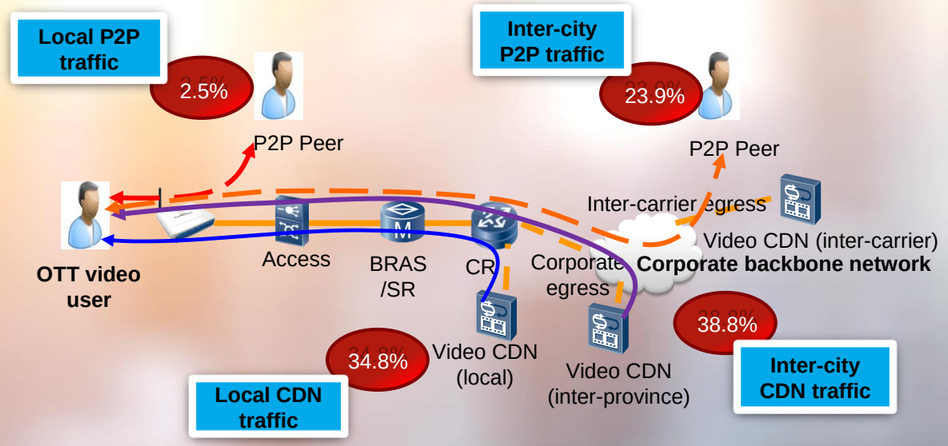
The native CDN and CDN+P2P streaming distribution modes have different impacts on carrier pipes.

(In addition to more distinctive traffic bursts, the CDN+P2P mode comes with extra upstream traffic.)





The CDN+P2P mode changes the traffic direction of metro networks. P2P traffic and inter-city CDN traffic under one viewing behavior together account for over 60% of the total traffic.



Technical White Paper

Every one wants superior video viewing experience. And what kind of factors determine video experience? Huawei iLab releases eight iLab service experience technical white papers that provide in-depth analysis on frequently asked questions. An excerpt is as follows:

OTT Video Interaction Closely Related to Initial Loading Time

The interaction process of OTT videos generally involves

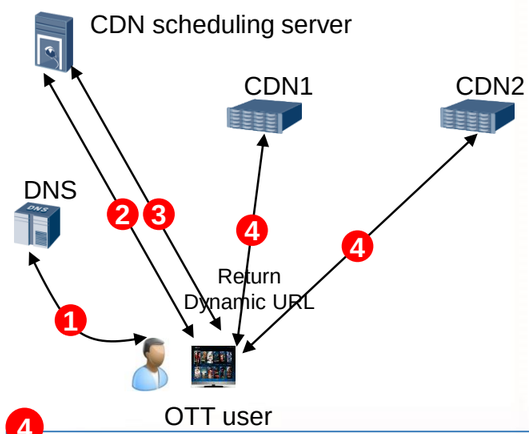
- DNS resolution
- Obtaining video segment information
- Obtaining video segment file addresses
- Downloading video segments

The following example illustrates the Youku video interaction process. During stages 1 through 3, the OTT terminal interacts with servers. At stage 4, initial buffering can be finished, and the OTT video can start playing only after the minimum data required for the playback to start is buffered.

1 The terminal accesses http://v.youku.com/v_show/id_*.htm to obtain the video web page framework and web page files, with the URL carrying the video ID. **The DNS server returns only the IP address of the CDN server (v.youku.com).**

2 The client requests http://v.youku.com/player/getPlayList/VideoIDS/*, with the URL carrying the video ID. The client interacts with the CDN server to obtain the video segment information, such as the segment file number, file type, file size, playing time, and video file type (SD, HD, or ultra HD).

3 The client requests the actual address of the video segment file from the CDN server, with information generated based on keys in the following URL: http://k.youku.com/player/getFlvPath/sid/*. The CDN server verifies the URL and generates a dynamic URL http://222.73.61.173/youku/AA/XX.mp4(flv) for the video file. The dynamic URL has a validity period of less than 8 hours and is generated based on the client's user-agent, request time, and random code. The dynamic URL varies for different users and repeated requests from the same user.



4 The client requests for video segment downloads on each CDN server based on the returned dynamic URLs.

The following formula is available: Initial loading time = Client-server interaction time + Time to buffer the minimum data required for the playback to start. Because the client-server interaction time is usually N x RTT (ms), the time to buffer the minimum data required for the playback to start mainly determines the initial loading time affecting user experience.

ICPs' Four CDN Scheduling Principles

Generally, ICPs follow four principles to determine policies for CDN resource scheduling: quality assurance, cost assurance, user distribution density, and corporate regional strategy.

The GSLB schedules CDN resources based on the capacity, physical bandwidth, and number of physical connections of every node, as well as the historical and current quality data between each user and each node, in order to achieve the best quality globally.

The GSLB schedules CDN resources based on the bandwidth costs, traffic costs, current bandwidth, and traffic usage of every node of all carriers in different areas, as well as user number predictions based on historical date, in order to ensure the lowest cost globally without affecting global quality.

User distribution density and corporate regional strategy: The GSLB differentially schedules CDN resources based on user areas, user levels, and terminal types.

Basic CDN Scheduling Algorithms

Different ICPs may have varied implementations of CDN scheduling algorithms. A CDN scheduling algorithm is a typical linear programming algorithm, which addresses the extreme values of linear objectives under linear constraints. To help understanding, the extreme value model can be simplified as follows:

Constraint 1: Cities A, B, and C have 500, 300, and 100 users, respectively.
Constraint 2: CDN nodes L, M, and N are available to provide services to these users globally.
Constraint 3: The capacities of the CDN nodes L, M, and N are 200, 300, and 400, respectively.
Constraint 4: The quality of services provided by the CDN nodes L, M, and N to users in cities A, B, and C are as follows:

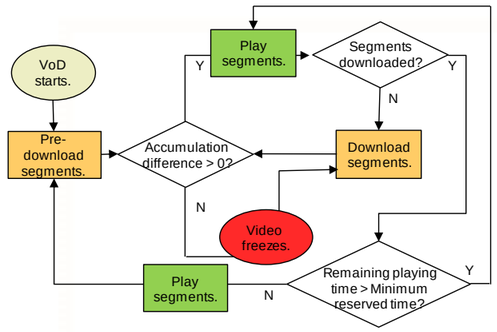
	L	M	N
A	90	80	70
B	80	70	90
C	70	80	90

Constraint 5: The total costs of services provided by the CDN nodes L, M, and N to users in cities A, B, and C are as follows:

	L	M	N
A	20	50	30
B	90	60	10
C	10	30	50

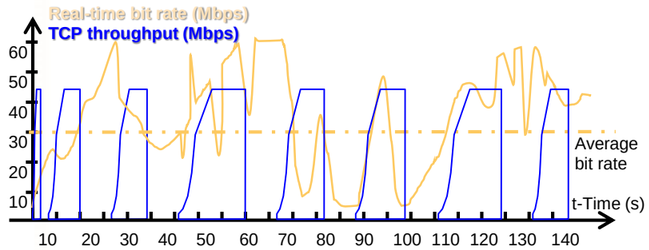
Question: How many users in each city should each node serve to achieve the best quality of service at the lowest cost?
This model shows that the CDN scheduling algorithms aim at an overall optimal solution. Quality and cost both considered, the service node scheduled to a user may not be the one physically closest to the user or with the best network quality.

Network KPIs Determine TCP Throughput and Affect KQIs of OTT Video Services



The OTT VOD services operated by mainstream Internet content providers are transmitted in the HAS architecture, which allows clients to use single TCP threads to download videos from a CDN. These services are streaming services that are progressively downloaded. Therefore, the download speed (TCP throughput) of the video segment directly affects the initial loading time, the number of video freezes, and the video freezing proportion.
Initial loading time = Client-server interaction time + Time to buffer the minimum data required for the playback to start
Because the client-server interaction time is usually N x RTT (ms), the time to buffer the minimum data required for the playback to start mainly determines the initial loading time affecting user experience. The higher the TCP throughput, the quicker the playback can start.

Mainstream OTT video files use a variable bit rate (VBR). The HAS architecture defines the OTT video segment download characteristics of pre-buffering, segment download, immediate download, and intermittent download. To ensure a viewing experience free from frame freezing, the amount of a video segment downloaded must be greater than the instantaneous bit rate of the playing time. Therefore, the TCP throughput directly affects the number of video freezes and the video freezing time, and accordingly the video freezing proportion.



For any given OTT video file, the distribution of instantaneous bit rates determines the lowest TCP throughput for the initial loading time, the number of video freezes, and the video freezing proportion to provide the best experience. The lowest TCP throughput required varies depending on the video files because they have different resolutions and instantaneous bit rate characteristics. The following example uses a group of Youku program sources measured in iLab. The TCP throughput for ensuring any movie in the program sources can be viewed free from frame freezing is as follows:

As indicated by the TCP throughput formula, the TCP throughput does not exceed the access bandwidth and is inversely proportional to the two-way delay and the result of root extraction of the packet loss rate.

	SD	HD	Ultra HD	1080P
Scenarios with ads (initial buffering of 8s)	450 kbps	832 kbps	3.8 Mbps	**Member-exclusive, no ads**
Scenarios without ads for members (initial buffering of a GOP)	800 kbps	1.6 Mbps	4.5 Mbps	8.3 Mbps

$$Throughput \le Min\left(\frac{WS}{RTT}, \frac{MSS}{RTT} * \frac{1}{\sqrt{p}}, \ MaxBW\right)$$

In other words, a lower access bandwidth results in a higher two-way delay or packet loss rate. If the two-way delay or packet loss rate exceeds a specific threshold, the TCP throughput will fail to meet best experience requirements, resulting in a long initial loading time, more video freezes, and higher video freezing proportion, and subsequently degraded user experience.

Network KPIs determine IPTV BTV KQIs.

During video playback, the bandwidth required from a terminal to a multicast replication point must be greater than the constant rate at which video streams are pushed.

Most IPTV BTV platforms owned by carriers send video streams at a constant video bit rate (BTV video source streams are generally transmitted at CBR). A BTV platform transmits only one copy of multicast traffic from the headend to a multicast replication point, which then replicates and sends the multicast traffic to terminals. To ensure that no garbled image or black screen occurs, the link bandwidth between the multicast replication point and terminal must be greater than the constant rate at which video streams are pushed.

If fast channel change (FCC) is deployed on an IPTV BTV platform, video streams in FCC acceleration are pushed at a rate 1.X times that of the video bit rate. Therefore, the link bandwidth from an iVSE-capable device to a terminal must be greater than the accelerated push rate.

If the link bandwidth is lower than the video bit rate, the image is damaged. The lower the bandwidth, the higher the image damage frequency.

The two-way delay determines the channel change delay.

When a user uses a remote control to change the channel being watched, the STB sends IGMP Leave and IGMP Join messages to a multicast replication point. After the STB receives the I frames in video streams of the new channel pushed from the multicast replication point, the STB plays images.

This process shows that the channel change delay approximates the two-way delay between the terminal and multicast replication point, if the communication time between the remote control and STB and the time for the multicast replication point to update the multicast group member entries are ignored. The two-way delay determines the channel change delay. IPTV experience shows that the channel change delay can be lower than 1s if the network delay is lower than 100 ms.

The packet loss rate determines the image damage frequency.

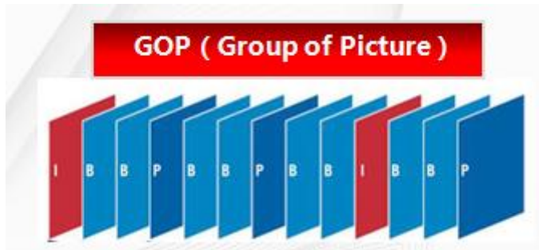
If a packet is lost and this packet contains I frames, the terminal displays a black screen transiently. If the lost packet contains only B or P frames, the image on the screen gets garbled here and there.

Initial Loading Time Varies Depending on the Terminal

Different OTT video terminals have different minimum buffered data requirements, which affect the time for buffering. The following table lists the minimum buffered data for different types of terminals.

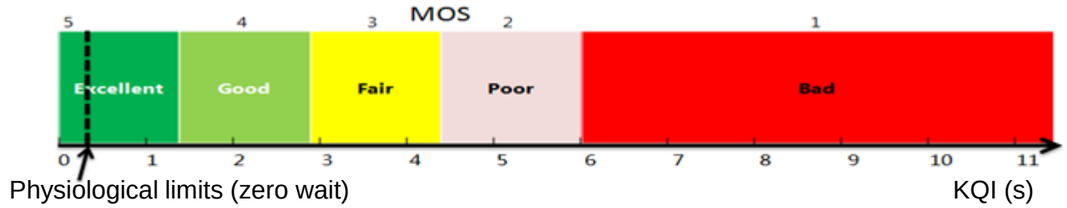
No.	Minimum Buffered Data	Typical Terminal
1	GOP or multiples of GOP	Flash Player on OTT web pages of Internet websites, such as Youku and Sohu
2	Amount of data required for playing a video for n seconds	VLC and Huawei A&S STB
3	First available I frame	Hisilicon's Android STB (The start of a video occupies the first I frame, and fast forward and rewind occupy the GOP at the play point.

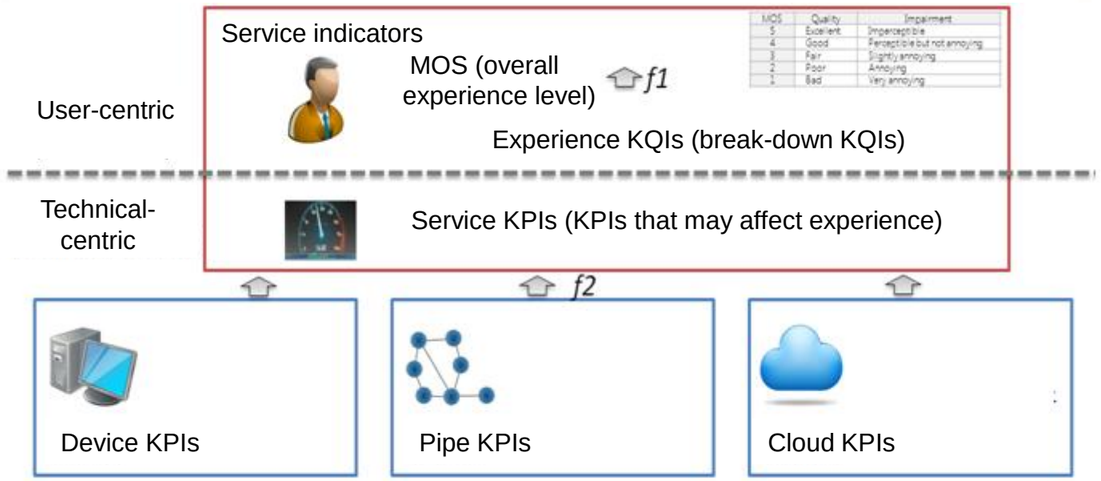
GOP is a group of successive pictures. In MPEG coding, pictures or frames, are categorized into three types: I, P, and B. An I frame is an intra coded frame that indicates the beginning of a GOP. A P frame is a predictive coded frame, and a B frame is a bipredictive coded frame. I frames contain the full image, whereas P and B frames record difference information relative to I frames. P and B frames cannot be decoded without I frames.



Traditional Internet service experience research

From experience to network: hierarchical mapping of indicators





Experience KQI: A user-centric indicator that is easy to understand for common users and helps the qualification and breakdown of user awareness

Service KPI: A technical-centric indicator that is used by O&M engineers for analysis of technical factors that may affect user awareness

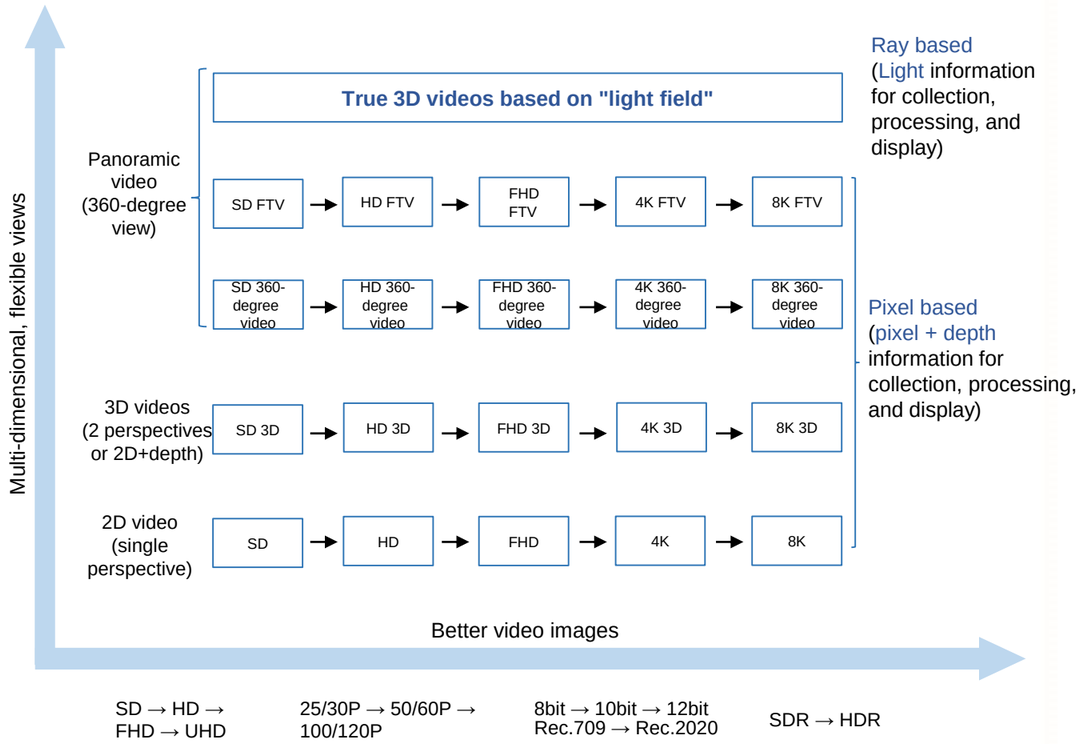
Definitions of KPIs, KQIs, and baselines for traditional Internet services

Service Type	KQI	Description	Measurement Method
File transfer	Download rate ratio	Ratio of the download rate obtained by the standard speed test software during access to a specified server to the bandwidth in the subscription package	Tested download rate/Download rate in the subscription package
	Upload rate ratio	Ratio of the upload rate obtained by the standard speed test software during access to a specified server to the upload rate in the subscription package	Tested upload rate/Upload rate in the subscription package
Online game	Operation response time	Period from the time a player clicks or taps to start the game to the time a client starts to respond, expressed in milliseconds	Game response time – Time a user clicks to start
	Operation freeze rate	Ratio of the number of freezes during a game to the total number of operations (terminal performance bottlenecks not considered), expressed in %	Number of freezes during a game/Total number of operations

Service Type	KQI	Description	Measurement Method
Web page browsing	Page response time	Period from the time a user enters a URL on the web page (with DNS resolution) to the time page loading starts. The time page loading starts when a title is displayed in the upper part of the browser.	Time when a title is displayed in the upper part of the browser – Time when the first DNS request is sent
	Initial screen display time	Period from the time a user enters a URL on the web page (with DNS resolution) to the time the data returned from the website is all displayed on the screen for the first time. Specifically, this KQI indicates the time required for displaying data in the upper part (1024 x 768 resolution) of the browser window	Time when the data is fully displayed in the upper part of the browser (1024 x 768 resolution) – Time when the first DNS request is sent
	Complete loading time	Period from the time a user enters a URL on the web page (with DNS resolution) to the time the complete page is loaded	Time when the last HTTP packet is arrived – Time when the first DNS request is sent

Service Type	QoE	KQI Baseline			E2E Network KPI Baseline		
					Bandwidth	Delay	PLR
Web page browsing	5	Page response time: < 0.6s	Initial screen display time: < 1.5s	Complete loading time: < 8s	>= 20 Mbps	< 20 ms	< 0.08%
	4	Page response time: 0.6s to 1.5s	Initial screen display time: 1.5s to 3s	Complete loading time: 8s to 18s	10 Mbps	< 15 ms	< 0.6%
	3	Page response time: 1.5s to 2s	Initial screen display time: 3s to 5s	Complete loading time: 18s to 30s	8 Mbps	< 45 ms	< 0.1%
File transfer	5	Download rate ratio: ≥ 100%	Upload rate ratio: ≥ 100%	/	50 Mbps	< 20 ms	< 0.01%
	4	Download rate ratio: 90% to 100%	Upload rate ratio: 90% to 100%	/	20 Mbps	< 25 ms	< 0.06%
	3	Download rate ratio: 70% to 90%	Upload rate ratio: 70% to 90%	/	10 Mbps	< 60 ms	< 0.07%
Online game	5	Operation response time: <0.4s	Operation freeze rate: 0%	/	>2.5 Mbps	< 60 ms	< 1%
	4	Operation response time: 0.4s to 0.7s	Operation freeze rate: 0% to 3%	/	>1.8 Mbps	< 70 ms	< 1%
	3	Operation response time: 0.7s to 0.9s	Operation freeze rate: 3% to 7%	/	>1.0 Mbps	< 80 ms	< 1%

Video — Images continuously improved, more views available, multi-dimensional



Network — Requirements — The exponentially increasing bandwidth requirements present huge challenges to network transmission.

Resolution	Video bit rate (H.265)			
	HD (<= 720p)	FHD (1920*1080)	4K (3840*2160)	8K (7680*4320)
"Light field" true 3D	50 Mbps to 10 Gbps**			
Multi-view FTV	10 Mbps to 40 Mbps*	15 Mbps to 70 Mbps*	50 Mbps to 300 Mbps*	>1 Gbps*
3D	1 Mbps to 2 Mbps	4 Mbps to 10 Mbps	30 Mbps to 50 Mbps	> 100 Mbps
2D	1 Mbps	2 Mbps to 5 Mbps	15 Mbps to 30 Mbps	> 100 Mbps

Remarks:

* Indicates that the bit rate is predicted based on the MVC coding technology; no commercial systems are currently available.

** Indicates that the bit rate is approximately calculated based on current capabilities using HEVC compression algorithm.

VR/AR

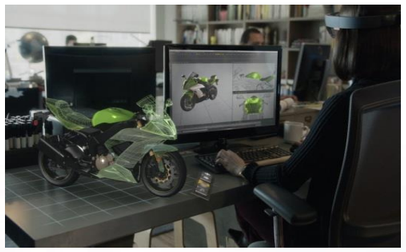
Office as a real-world environment, skeleton as a virtual object

- **Virtual Reality (VR**) is a 3-dimensional virtual world that uses computer simulation to artificially create sensory experiences such as sight, touch, and hearing. It offers real-world experience, allowing users to observe objects in real time and without bounds. This, however, requires **concurrent output of a huge amount of sight, touch, and hearing information as well as real-time interaction. Therefore, high-speed transmission becomes a challenge.**
- **Augmented Reality (AR)**, also known as mixed reality, uses information integration technologies to integrate virtual information into a real-world environment. In this way, virtual objects and the real-world environment are overlapped in the same image or space.

Interaction difference:

- VR device: VR devices are used for interaction between users and a virtual environment. The most frequently used VR devices are the position tracker, data glove (like 5DT), capture system, and data mask.
- AR device: Because AR is a combination of real-world and virtual environments, a camera is required. Microsoft HoloLens is an example of an AR device, which can display image shoot using a camera and allow interaction between real-world and virtual images.

VR device: Oculus Rift



AR device: HoloLens

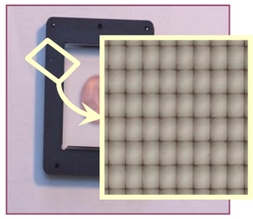
Source: Light field materials from the Multimedia Communication Lab in Xidian University

Light field-based Future 3D videos

The light field describes the amount of light flowing in every direction through every point in space. We can use the light information recorded to generate video images for digital focus in any perspective and from any distance.

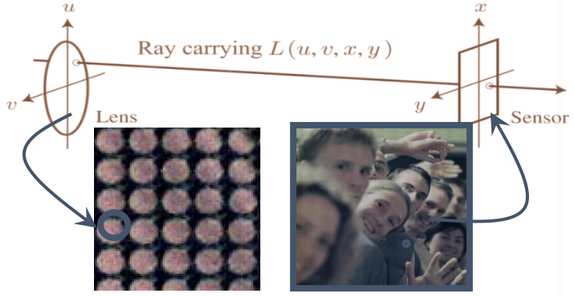
Demo of light field

Camera arrays



Microlens array

Light field capture device



Light field capture picture

If the sampling density of lights on a plane of focus is 300*300, the counterpart on a camera plane is also 300*300. The amount of the light field data before compression is 190G bit (**8bit x 3 x 300 x 300 x 300 x 300)**.

The **Refocus** and **Free Viewpoint** features of the light field provide a better restoration of 3D scenarios and better 3D experience that is closer to human eyes. Watching traditional 3D movies for a long time or using traditional VR devices leads to strong visual fatigue, reducing user experience. This problem can be addressed in the light field.

- **Refocus:** While observing a 3D world with human eyes, the focused objects are sharply imaged in human eyes. The farther away from the plane of focus, the more blurred the images. While observing objects with different depths with human eyes, the sharpness of images varies accordingly.
- **Free Viewpoint:** Object observation with human eyes from different angles comes with different parts of the object.

Source: Light field materials from the Multimedia Communication Lab in Xidian University

2015 Big events of Huawei iLab

February, 2015

Completed service experience baselines for **traditional base services**, including Web, OTT, download, upload, and online gaming.

May, 2015

Worked out **online game characteristics** and an experience analysis report.

August, 2015

Cooperated with the Quality Department and incorporated 12 KQIs into TL 9000.

October, 2015

1. Worked out the mobile bearer U-vMOS network requirement baselines.
2. Worked out experience benchmarking reports for five cities in China: Beijing, Shenzhen, Xi'an, Nanjing, and Chengdu.
3. Provided OTT video, bandwidth speed test, web page test algorithm, and SDK, and incorporated SpeedPro, taking the first step to make experience measurement public.

December, 2015

1. Confirmed cooperation intentions with Brillview, Zafaco, and Tencent and will sign the MOU in the near future.
2. China MIIT's industry standard, *Broadband Internet Service and Network Quality Evaluation Methods*, was approved.
3. Huawei unveiled *the 2015 Big Data Report on China's Internet Videos*.

April, 2015

1. Worked out reports on **smartphone games** characteristics and experience analysis.
2. Developed **KQIWatch**, the prototype of the experience evaluation tool, which is equipped with lightweight functions of collecting experience indicators, service indicators, and network indicators of traditional Internet services. This makes collection of experience indicators possible and was successfully used in the experience evaluation project for **Beijing Mobile**.

July, 2015

Output the service experience bottleneck analysis and optimization baselines. The experience evaluation project for Beijing Mobile expanded the market by **70 million CNY.**

September, 2015

1. Unveiled U-vMOS, the video experience evaluation system, at the UBBF.
2. Partnered with Ovum to unveil BES, the global broadband experience scorecard.
3. Produced the Internet data crawler that supports bit data parsing and analysis on massive Internet service characteristics.

November, 2015

The U-vMOS UDP algorithm standard that is jointly developed with Network Research Department was approved at an ITU-T meeting, and the UDP H.264 algorithm has been provided in SDK mode.

Fixed network U-vMOS UDP algorithm standards approved by ITU-T

At the ITU-T SG12 Q13/14 meeting held in Berlin, Germany in November 2015, the guidelines for U-vMOS UDP algorithm standard (G.OM_HEVC, an HEVC quality evaluation model in BTV UDP scenarios) proposed by Huawei fixed network line were approved by all the participating parties and well acclaimed by various vendors for further cooperation.

This project plans to work out the first HEVC/H.265 related video experience measurement standard (video MOS) in the industry in 2016.

Huawei iLab's user experience standards for typical services incorporated into the indicator set at TL 9000 forum

In 2015, Huawei iLab, with a focus on Internet service experience, developed the customer-aware service experience indicator model and baseline and continuously explored the applications of experience indicators to optimize the experience data operation mechanism. This effectively addresses the lack of a model indicating the relationships between experience indicators and network quality and makes customers' service experience measurable. The service experience standard based on customers and networks was incorporated into TL 9000.

Broadband Internet Service and Network Evaluation Methods approved by China's communication industry

At the fourth meeting of CCSA TC1 WG1, held on December 16, 2015, the document *Broadband Internet Service and Network Quality Evaluation Methods* – which is jointly worked out by Huawei fixed network iLab, Huawei Research Dept, China Unicom, and China Telecom - was well acclaimed by all the participating members and has entered the approval stage. This signifies the remarkable progress that Huawei Fixed Network iLab has made in experience standardization in China's communication industry.

Huawei iLab is committed to research on the best user experience. This standard aims to provide consistent measurement methods for service quality evaluation from ISP/ICP, covering web page browsing, broadband speed test, mobile speed test, and OTT videos. It has defined the experience evaluation architecture, service measurement indicators, measurement methods, measurement solutions, and sampling principles.