White Paper on the VR-Oriented Bearer Network Requirement(2016)

Issue 01

Date 2016-09-15

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

Huawei Technologies Co., Ltd.

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

Website: http://www.huawei.com

Email: support@huawei.com

Contents

1 VR Service and Application Overview

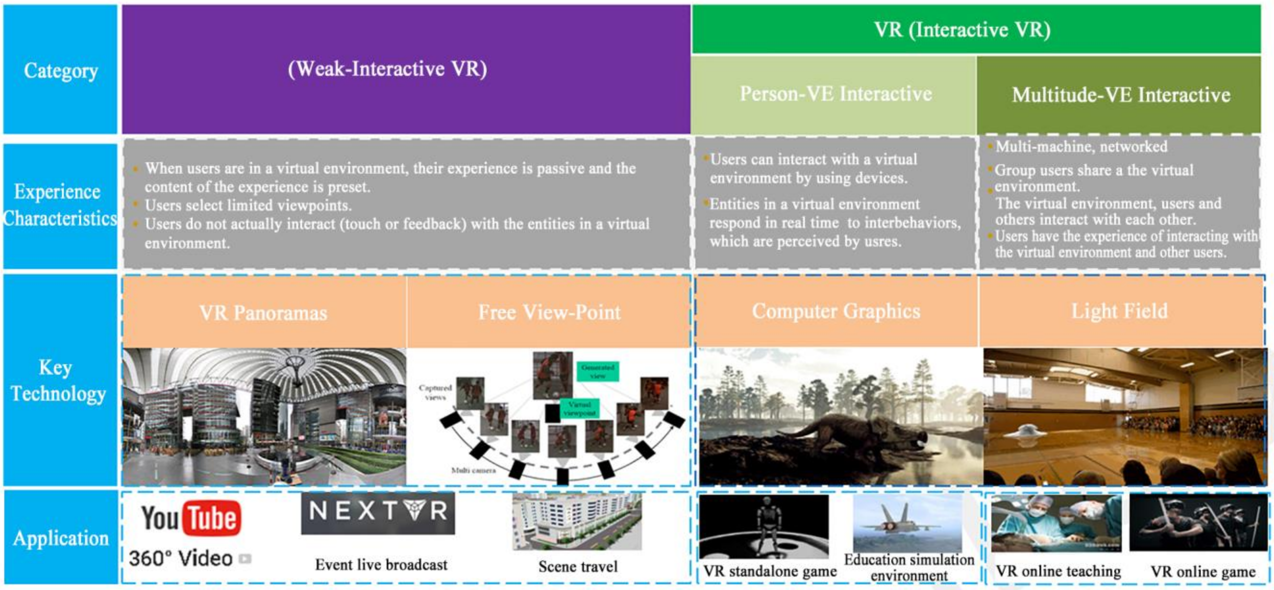
Virtual reality (VR) is the next-generation Internet and next-generation computing platform. VR is publicly recognized as a virtual environment created by computers. The environment replicates a real world or creates an imaginary setting where users interact in real time with the virtual space.

VR is multi-sensory. According to conceptual models proposed by J.J.Gibson, the perceptual systems of humans include the sight, hearing, touch, smell/taste, and direction. VR is supposed to satisfy all of these senses of users. In MPEG seminars, VR is considered as a new media type, different from video and audio.

Based on this definition, VR involves a variety of key technologies and applications. For example, the key technologies include:

- 360 Panoramic Video;
- Free viewpoint;
- Computer Graphics;
- Light Field and so on.

Based on the above key technologies, various applications are generated from VR, such as online VOD & BTV based on the 360 panoramic video technology and free viewpoint technology and VR standalone games, and computer graphics-based VR console games, online games, and VR simulated environments.



These applications target different market scenarios. For example, Goldman Sachs classifies the VR market domain into games, event live broadcasts, entertainment videos, healthcare, real estate, retail, education, engineering, and military affairs.

Three elements of VR are as follows:

- Spatial- Virtual environment information perceived by users is spatial and contains mass information.
- Interact-Users can conduct information interaction with spatial data in virtual environment and other users. Information forms connection and flow among users.
- Real-Time-Information interaction of users in virtual environment is real-time, which requires real-time information connection.

According to the three elements of VR, any VR application involves real-time connection and flow of mass information when it comes to online network applications, and brings new challenges to the network architecture.

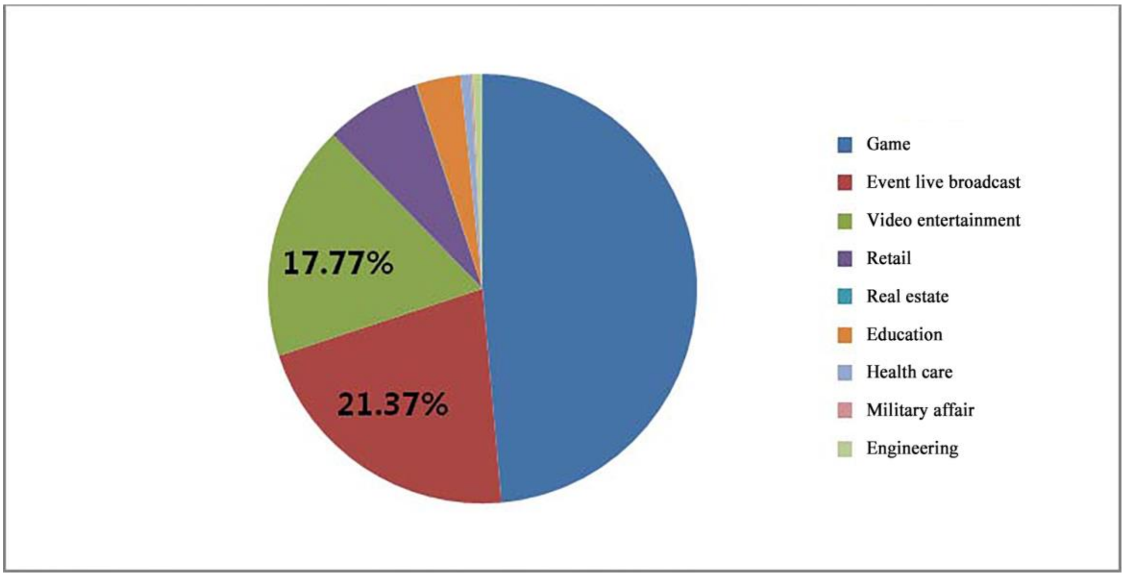
2 VR 360 Video Is the First Prosperous Online VR Service

2.1 Development Trend of the VR 360 Video Industry

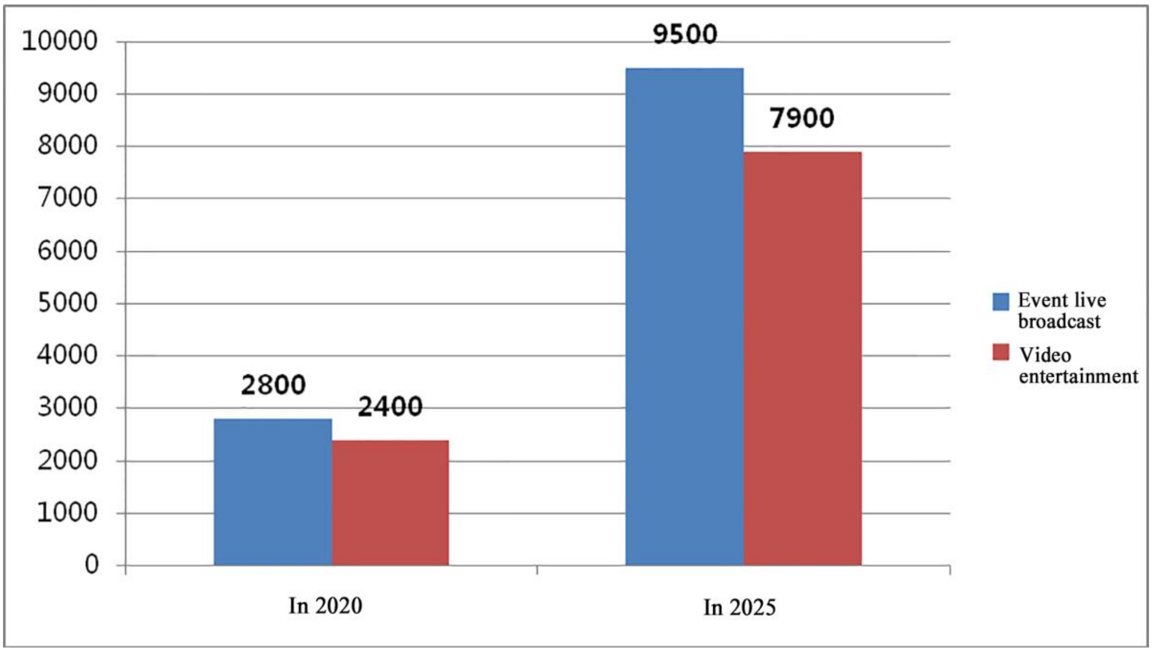
VR 360 Video provides a physical space field of vision surrounded by the 360 °horizontal direction (longitude) and the 180 °vertical direction (latitude) at the observer's physical location. Users can switch viewpoints by changing the position of their heads or by input devices such as a mouse and remote control so as to bring about immersive experience.

Combining analysis on industrial elements such as users, technologies, hardware, content and standards, we think that among numerous key technologies and applications of VR, VR 360 video based on the panorama technology will become the first prosperous online VR application.

According to the VR/AR industry report published by Goldman Sachs in 2016, VR event live broadcasts and VR video entertainment based on the 360 panorama technology will own 52,000,000 users in 2020, including 24,000,000 for event live broadcasts and 28,000,000 for video entertainment, which accounts for 40% of all expected users (130,000,000) of VR application domain. In 2025, users of VR 360 video will be 174,000,000, including 95,000,000 for event live broadcasts and 79,000,000 for video entertainment.



Unit: 10,000



Currently, driven by the Internet industry's layout, content sources and user groups of the VR 360 video have taken shape:

Content sources of the online VR 360 video increase rapidly for the latest year. YouTube sets the 360 video area and the number of newly uploaded VR 360 videos is over 8000; the market video number of Oculus 360 Video which provides content to Samsung Gear and Oculus CV1 is over 1000; Video is a content platform focusing on 360 videos, which provides high-quality 360 video online content to HTC Vive, Oculus and Samsung Gear. Among them, full-view 360 videos of 4K resolution is over 400, accounting for 70% of the total number; domestic traditional Internet video companies including Youku, Letv and iQIYI also set the 360 video area, focusing on establishment of content ecology. Most of the video sources are home-made variety shows, with the number of several hundred; online services of Next VR includes VR 360 live broadcasts/playbacks, VR 360 films and documentaries on demand. It has carried out high-quality VR 360 live broadcasts successfully for well-known matches such as NBA, U.S. Open (golf) and International Champions Cup.

VR 360 video consumption also grows rapidly. The user base and clicks of VR videos, especially hot videos, are also huge. Top N 360 videos at YouTube have average daily clicks of 205,000. Hot 360 videos of the Youku VR channel have average daily clicks of 40,000. Samsung Gear also has 1000,000 monthly active users.

2.2 Trend Analysis of VR 360 Video User Behavior

2.2.1 Prediction of User Penetration Rates

The user penetration rate of VR 360 video services and its traffic trend are crucial to analyzing their influences on the bearer network architecture.

The *Internet Trends* 2016 report issued by Mary Meeker-the Queen of the Internet showed that there were 3 billion global internet users at the end of 2015, but the growth rate showed a stable down trend for a consecutive of four years between 10% and 9%. Combining the historical data with the analysis data reported by a consulting company, the trend data for global internet users with some common methods are shown as follows:

Year	**2015**	**2018**	**2020**	**2025**
Linear trend prediction (Between 9% and 8%)	3 billion	3.8 billion	4.4 billion	5.8 billion
Binomial trend prediction	3 billion	3.8 billion	4.4 billion	6.2 billion
Analyst report	/	3.6 billion (eMarketer)	/	4.7 billion (Microsoft)

Combined with the above data, the predicted values of global internet user number for us in 2020 and 2025 are 4.4 billion and 6 billion respectively. According to the VR 360 video user number prediction given by the Goldman Sachs report, the user penetration rate at the scheduled time is shown as follows:

Application Service Penetration Rate	**2020**	**2025**
VR entertainment video	0.55%	1.32%
VR event live broadcast	0.65%	1.58%

2.2.2 Prediction of User Viewing Habits and Average Traffic

As previously mentioned, due to terminal and content experience issues, the playing time for a user every time is hardly more than 20 minutes with existing VR technology in 2016. During the transition of VR 360 video popularization, the industry will adopt a compromise propose combining traditional terminal with complementary HMD. With the improvement of terminal and content experience, the playing time for a user every time continuously increases. It is predicted that the longest playing time for a user every time is up to 60 minutes in 2020, and above 120 minutes in 2025, which is similar to the current traditional viewing habit.

1. Entertainment Video:

   The *China VR User Behavior Survey (2016)* indicates that the daily average usage time of current VR user is 34 minutes, of which the playing time for VR 360 video is about 10 minutes. However, the Goldman Sachs report indicated that the potential user group of VR entertainment video is similar to the today's Netflix online video user. That also is the target market of the VR video content in a long run. The VR entertainment video viewing level in 2025 shall reach that of Netflix traditional video user. However, the TDG Research in 2015 indicated that the daily average playing time of the subscribers of Netflix on the global scale is up to 93 minutes, which is what we predicted the daily average playing time of the VR 360 entertainment video user in 2025.

   In view of the terminal characteristics, we think that the VR 360 entertainment video is generally used in the living room. The usage time ratio between the evening leisure time period (19:00-23:00) and other time period for users is 8:2. That is, an average playing time of evening leisure time period an hour is 20% of the daily average playing time. According to the peak load shifting effect of a user group using service at the same time, the overall traffic generated for using the service is equally distributed within an hour. Based on the presumption, we can get the user average traffic for watching entertainment video approximately.

	2016	2020	2025
Daily average playing time (minute)	10	52	93
Proportion of evening leisure time period (19:00-23:00)	80%	80%	80%
During evening leisure time period (19:00-23:00), service time generated by a user in an hour (minute)	2	10.4	18.6
Traffic of evening leisure time period (19:00-23:00) B:VR 360 video bitrate N1: network VR entertainment video user number	2*60*B*N1/3600	10.4*60*B*N1/3600	18.6*60*B*N1/3600

2. Event Live Broadcast:

According to the Goldman Sachs report in 2016, the users will watch 2 events live per year at first. As the contents become rich, the viewing number will be increased gradually, which is about 4 in 2025. The duration of a concert is mainly between 1.5 hours and 2 hours; the live duration of a NBA basketball game is between 2 hours and 2.5 hours; the live duration of a football game is about 2 hours. In conclusion, an event duration is 2 hours. Considering that the ratings of the event live broadcast have the extremely obvious tidal effect, the daily average playing time cannot reasonably reflect the user habit of the event live broadcast, the audience rating shall be used to measure its effects on a network. Audience rating refers to the percentage of the number of people watch a certain program in a certain period of time from the total number of audience. We think that the live events can be classified into major events (such as Olympics Opening Ceremony, 70th Anniversary of the Victory of the Chinese People's War of Resistance), hot events (such as American Presidential Election), and general events (such as concert and variety show that will be followed by users with specific interests or preferences) according to the influence area and attention. We will obtain the live audience rating data of these events in history for reference on analysis. Meanwhile, due to common sense, we think that the audience rating of an event in a certain grade will not change with the changes of the time. The factors that will impact event live broadcast traffic in different ages mainly depend on VR 360 video bitrate and user total number of event live broadcast.

	Major Event	Hot Event	General Event
Audience rating R	8%-15%	2%-8%	0.5%-2%
Traffic during event live broadcast B:VR 360 video bitrate N2: network VR event live broadcast user number	N2*R*B		

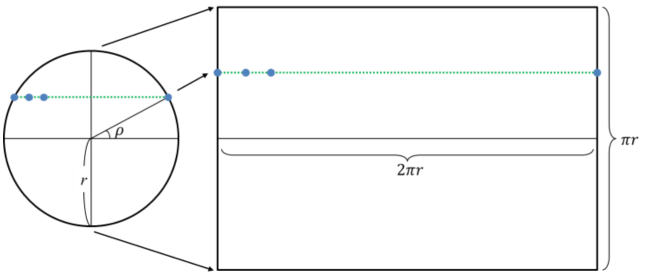
3. Overlay:

	2016	**2020**	**2025**
Live event during evening leisure time period (19:00-23:00)	2*60*B*N1/3600+ N2*R*B	10.4*60*B*N1/3600+ N2*R*B	18.6*60*B*N1/3600+ N2*R*B
Live event during other time period	N2*R*B(Compared with the evening leisure time period, the on-demand traffic in other time period is relatively small, which can be neglected)		

3 Service Principle and Network Requirements of VR 360 Video

3.1 Full Perspectives and FOV

The vision of users in the virtual environment is considered a space ball, which unfolds 360 ° horizontally and 180 °vertically. With a terminal, users receive part of the spherical data though either eye, the extent to which depends on the Field of View (FOV) provided by the terminal.



If FOV is 90 °, 1/8 of the spherical data is received; if FOV is 120 °, 2/9 of the spherical data is received. This is not applicable for VR on traditional terminals, such as televisions, pads and phones, where the angle of view is far smaller than 90 °.

3.2 Image Quality and Interactive Experience

The issues for VR experience focus on perception and physiology. The physiological issues are an imperative in the industry for the widespread use of VR. The players in the market generally encounter the four physiological issues as follows:

Issue 1: Quality of visual information. As known by all, the visual fatigue caused by low image quality can lead to dizziness. The industry strives to optimize the quality of VR 360 video, to upgrade the resolution ratio and the image quality over recent years.

Issue 2: Head movement and motion-to-photons latency (MTP). The mainstream of industry believes that the latency should not exceed 20 ms, otherwise it should cause dizziness. Leading VR terminal manufacturers, such as Oculus and HTC Vive, have started with sense tracking components, display screen technique and GPU and managed to reduce the localization of motion-to-photons latency to 20ms by promoting the performance of end-to-end software and hardware.

Issue 3: The conflict of motion perception. Lack of the output of motion feedback may cause mismatching of motion and the visual information seen, which leads further to dizziness. To solve this issue, it requires multi-perceptibility of abundant VR terminals from the industry to provide fusion abilities including vision, hearing, touch and motion feedback and to develop the function of VR new media to full extent.

Issue 4: Vergence-accommodation conflict, or focusing conflict, occurs in a display terminal that utilizes the theory of binocular parallax. Because no depth data is provided by the light emitted from the screen, the focus of the eyes is on the screen. The focus adjustment of the eyes mismatches with the visual depth, which causes dizziness. This experience issue requires a new technology, which records and restores the intensity and angle of the light emitted from a spot in the three-dimensional space by using the record of the light field and the projection technology. This technology will mature in the future.

This section will demonstrate on the two issues which the industry have focusing for breakthrough, which refers to the image quality experience corresponding to the quality of visual information and the interaction experience corresponding to the motion-to-photons latency.

1. Image quality experience:

   Due to the difference between the full perspectives and FOV in the virtual reality, the traditional description of resolution ratio for OTT video corresponds to spherical full perspective resolution ratio for VR 360 video. What determines the image quality experience of VR 360 is the monocular resolution ratio (FOV resolution ratio), which can be converted into pixels per degree (PPD) visible in each angle in the FOV area. The higher the PPD value is, the higher the PPD of the field of view is, and the better of the image quality experience will be. The maximum resoluble PPD for users of normal eyesight is 60. Average person cannot tell the distances among pixel points if the PPD exceeds 60.

   For example, the online VR 360 video from YouTube with 4K resolution employs H.264 coding at an average code rate of 20 Mbps, which is of the highest grade. However, the actual monocular visible resolution of 4K resolution in spherical full perspective is only 960*960, which has only 10 pixels in each degree in the corresponding 90 °field of view. This value is far beneath 60 PPD required by the retina of normal eyesight. The actual video experience is worse than SD video on traditional TV, PC and Pad.

Screen Type	Screen Size (inch)	Viewing Distance (meter)	Width (meter)	Height (meter)	Horizontal Resolution	Vertical Resolution	PPD	FOV
TV	60	1.5	0.98	0.55	360	240	10	36
PC	24	0.6	0.39	0.22	360	240	10	36
Pad	10	0.25	0.16	0.09	360	240	10	36

   Drawn from the above example, because an immersion terminal (HMD) of VR has a field of view larger than traditional terminals (TV/PC/Pad/Phone), a VR 360 video requires a higher monocular resolution ratio and full perspective resolution ratio to meet the PPD requirement for a competent image experience. The full perspective 4K resolution ratio is far behind the required video quality. It is a necessity to increase the resolution ratio to over 8K. Take FOV = 90 as an example, when the full perspective resolution ratio reaches 8K, the monocular resolution ratio is 1920 x 1920, PPD = 22;

when the full perspective resolution ratio rises to 12K, the monocular resolution ratio is 1920 x 1920, PPD increases to 32. We will demonstrate on the evolution route of VR 360 video image quality experience later.

2. Interactive experience:

   According to an academic research, VR applications are divided into weak-interactive VR and interactive VR by the interaction experience angle between a user and virtual environment (VE). VR 360 video is weak-interactive VR, where users passively experience pre-filmed contents in a virtual environment. Users alter viewpoints by swiveling but cannot engage in substantial interaction with the virtual environment.

   Accordingly, the interaction experience of VR 360 video is mainly reflected in head movement and motion-to-photons latency (MTP). Mainstream of industry believes that while using immersion terminals, the MTP cannot exceed 20ms, otherwise it should cause dizziness. In other words, while users alter viewpoints, the global MTP of terminal, network and cloud processing should guarantee the uniformity of head movement and FOV image alteration. The image update latency should not pass 20ms, nor should whole/part of the view loses image information.

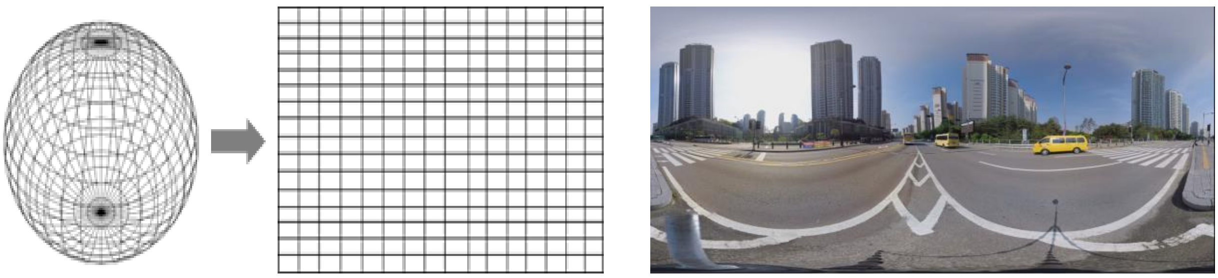
3.3 Projection and Coding Techniques

Projection and coding techniques determine the format in which VR 360 videos are produced and organized, and the amount of medium information contained. This is critical for quantifying network requirements for a certain user experience.

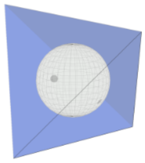
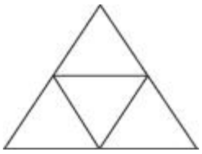
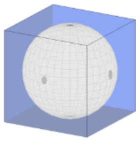
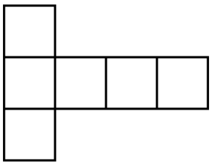
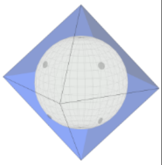
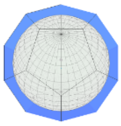
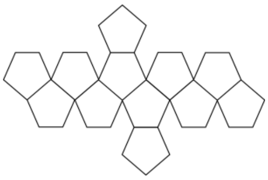
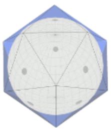
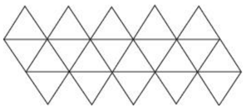
1. Projection technique

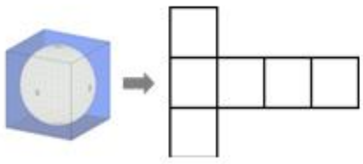
   VR 360 video needs to alter the space ball information into two-dimensional media format, which requires projection technique that traditional videos never employ.

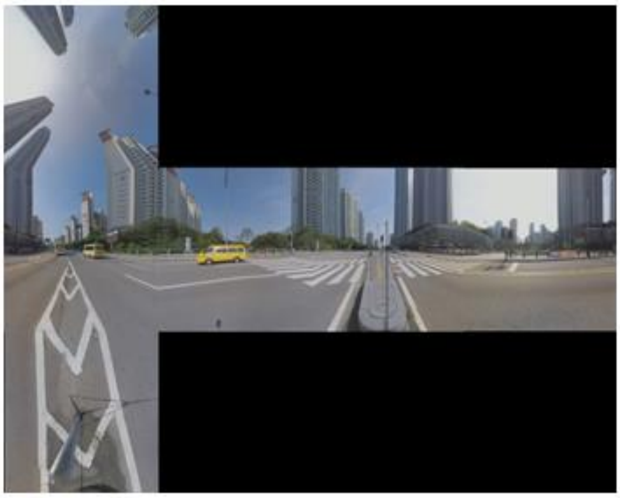
   Currently in the mainstream format of VR 360 video, equirectangular projection (ERP), image quality may manifest distortion, and the compression efficiency is within bottleneck. This approach of projection employs a typical idea of projecting meridian and parallel and unfolds the sphere into two-dimensional rectangle. The orthogonality of meridian and parallel for isometrical projection is 90 °with no angular distortion but maximum area distortion. The angle stays stable mainly due to the increase of the area. The projection of the equator of the sphere manifests small degree of quality distortion which increases towards the two ends. While the two ends unfold and the angle stays the same depending on the area increase, more ineffective redundant pixels are introduced, which causes low efficiency in video coding and compression. YouTube, Oculus and Samsung Gear abroad and domestic Youku and iQiYi all adopt VR 360 media files produced by this projection.



Platonic solid projection (PSP) with its low degree of distortion and high efficiency of compression becomes the new direction of industry. It utilizes another typical map projection idea, which divides the sphere into numbers of spherical trapezoids based on meridians and parallels with the same difference of longitude and latitude and project to certain polyhedron. Polyhedron here can be tetrahedron, cube, pyramid and dodecahedron. Each trapezoid is projected individually, which leads to low degree of distortion. Samsung submitted the proposal on PSP projection on the MPEG meeting in May, 2016.

Projections	**3D Model**	**2D Projection**	**Vertexes**	**Area Ratio**	
				vs. Sphere	**vs. ERP**
Tetrahedron (4 faces)			4	3.31x	2.11x
Cube (6 faces)			8	1.91x	1.22x
Octahedron (8 faces)			6	1.65x	1.05x
Dodecahedron (12 faces)			20	1.32x	0.84x
Icosahedron (20 faces)			12	1.21x	0.77x





2. Coding Technique

   VR 360 video can be compressed with ordinary video coding technologies. H.264 is the most applied video coding technology, and HEVC and VP9 are the industry-recognized coding technologies for the next generation. According to the conclusion of a recognized test, the compression efficiency of HEVC and VP9 is higher than that of H.264 by 30%. The latest research of MPEG and other standards organizations suggests that the compression efficiency of the next-generation coding (H.266) is 30% higher than that of HEVC.

   The VR 360 video with effect of field depth (3D), is synthesized with two images with binocular parallax in left and right eyes, so as to form a three-dimensional (3D) effect. It is reflected on the media format that the two images corresponding in left and right eyes are coded into same frame, and possibly in lateral arrangement or vertical arrangement. For the information quantity without compression, the 3D-effect VR 360 is twice than the 2D-effect one. As the contents in left-right eyes have higher correlation, the compression efficiency can be further enhanced to achieve the same image quality. The industry test results show that, with the same version of coding technologies, the compression efficiency of 3D-effect VR 360 video can be enhanced 25% further above the 2D-effect VR 360 video.

3.4 Network Transmission Technology Routes

The online transmission of VR 360 video includes two main technology routes: omni-perspective transmission scheme and FOV transmission scheme.

- **Omni-perspective Transmission Scheme**

  The omni-perspective transmission scheme is to transmit 360 °surround image to the terminal; when a user head rotates to switch the image, all the processing is finished in the local terminal. Under the same situation of monocular visual resolution, for such reasons as frame rate, bit depth, 360 degrees, the code rate of VR panoramic video is much larger than that of normal planar video; the former generally is 5 to 10 times larger than the latter. The extreme panoramic VR video perfection of monocular 8K, for example, its bandwidth is required to achieve 5G during watching, which can be a great challenge for the network speed and also increase the costs greatly.

While the whole panoramic video is in 360 °, actually viewers can only see the current vision part during their viewing; thus, the unseen part occupies the network bandwidth but is not really utilized, causing larger waste of network resources. Against this situation, this industry sector proposed an FOV transmission scheme for differentially transmitting VR video on the basis of visual angle.

- **FOV Transmission Scheme**

However, the FOV transmission scheme mainly for transmitting the visible images in the current visual angle, generally divides a 360 °panoramic vision into multiple visual angles, and generates one file respectively for each visual angle, only containing the visual information of the high resolution within visual angle and the low resolution in surrounding parts; the terminal requests the server for the corresponding visual-angle file according to the user's current visual-angle posture position. When the visual angle changes by user head rotation, the terminal requests the server for the visual-angle file corresponding to new visual angle.

Among the FOV transmission schemes published in Facebook, totally 30 visual angles are divided, and each visual-angle file is only 20% of the original file; the transmission code-rate is accordingly only 20% of the original, greatly reducing the bandwidth requirements of viewing VR video, and improving the effective bandwidth utilization. But this scheme also has some shortcomings, such as the total sum of video file size for all the visual angles is 6 times of the original file, occupying more storage space in the server; however comparatively, the bandwidth resources is more valuable.

The following will give more detailed elaboration on these two schemes.

3.4.1 Omni-perspective Transmission Scheme

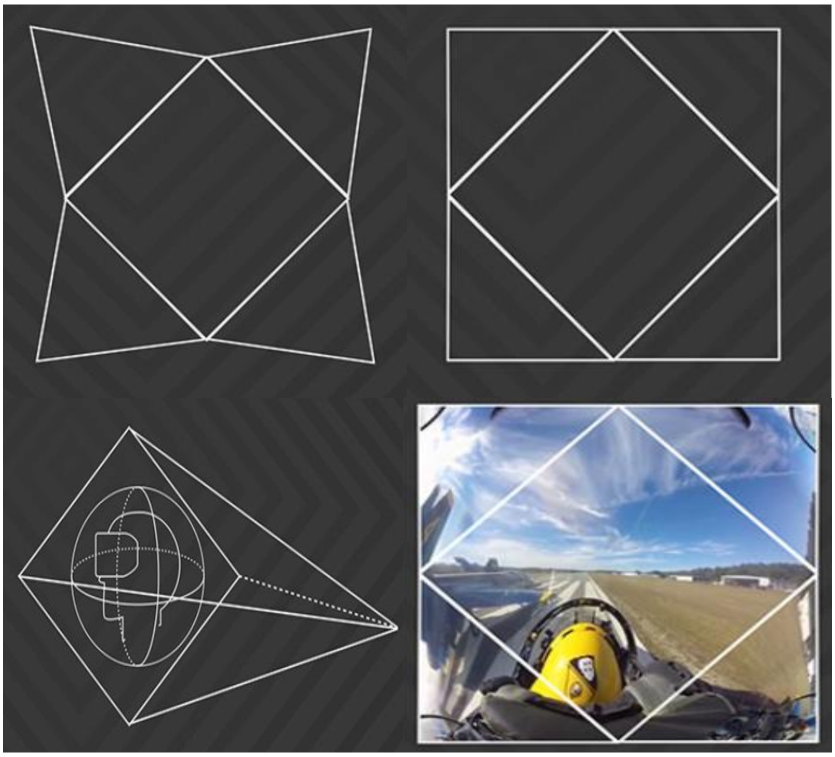
In the omni-perspective transmission scheme, a frame of the data received by a terminal contains all the visual-angle information for the space ball visible to a user. The terminal will process the interactive signals of changes to visual angles; the terminal extracts FOV information from the cached frames according to visual-angle information. After the FOV information is corrected and restored in the player, the users will obtain the visual information in a normal angle. Therefore, the 20 ms required by interaction experience is ensured by the terminal, not involving the network delay and cloud-end delay. This scheme requires a higher bandwidth speed but a lower delay. This is a transmission scheme of "bandwidth exchanged for delay". This transmission scheme directly utilize the current mainstream video transmission technologies, such as MPEG.DASH, HAS, HLS, and HPD; it adds a function of projecting and restoring FOV information from omni-perspective frames in a terminal player; the details of its flow will not be discussed in this paper.

3.4.2 FOV Transmission Scheme

In the FOV transmission scheme, one frame of data received by the terminal never contain all the visual-angle information without distinction of the space balls, but structure the corresponding frame data according to the user's visual-angle posture; one frame of data only contain the visual information of the part ≥ view-field angle; while the terminal needs to judge the posture position of visual angle caused by user head rotation to send the interactive signals to the cloud end, and to request the frame data corresponding to new postures. So the 20ms required by interaction experience contains not only the time delay of terminal processing and also the time delay of network transmission and cloud processing. This scheme requires lower bandwidth speed but higher time-delay (E2E<20ms), belonging to a transmission scheme of "time-delay exchanged for bandwidth". This transmission scheme can directly utilize the current mainstream video transmission technologies, such as MPEG.DASH, HAS, HLS and HPD; but it accordingly requires the generating mechanism of media files, the processing mechanism of cloud and terminal, thus to form a peculiar flow. The following article gives more elaboration on its detailed flow.

1. Posture definition: define the visual-angle posture position of enumerated user in space balls, and number it from **1~N**; each No. corresponds to a visual-angle range, and the visual angle range can just equal FOV and also be greater than FOV.
2. Media generation: generate N media files according to the number of a visual-angle posture position, and store them in a cloud server. Prepare the media description file (MPD) on the cloud server.
3. When playing the VR 360 video, the terminal first requests the cloud for the access to media description file (MPD), to obtain the relationship of a user's visual-angle posture position and media files.
4. The terminal requests the media file **i** and the time-point **t0** of VR 360 video playing start, according to user's current visual-angle posture position. After cloud receives the request, it addresses the corresponding time **t0** of the media file to begin transmission. Terminal begins playing at the time **b0** of receiving the minimum buffer data volume supporting playing, and continues to request the cloud for subsequent media file content.
5. If the user changes its visual angle at the time **t1**, for each agreed visual-angle change degree **Φ**, the terminal identifies out the corresponding visual-angle posture position as **j**, and the Position **i** and Position **j** are continuous adjacent two positions. Terminal requests the corresponding media file **j** and the time-point **t1+Δt** of VR 360 video playing beginning.

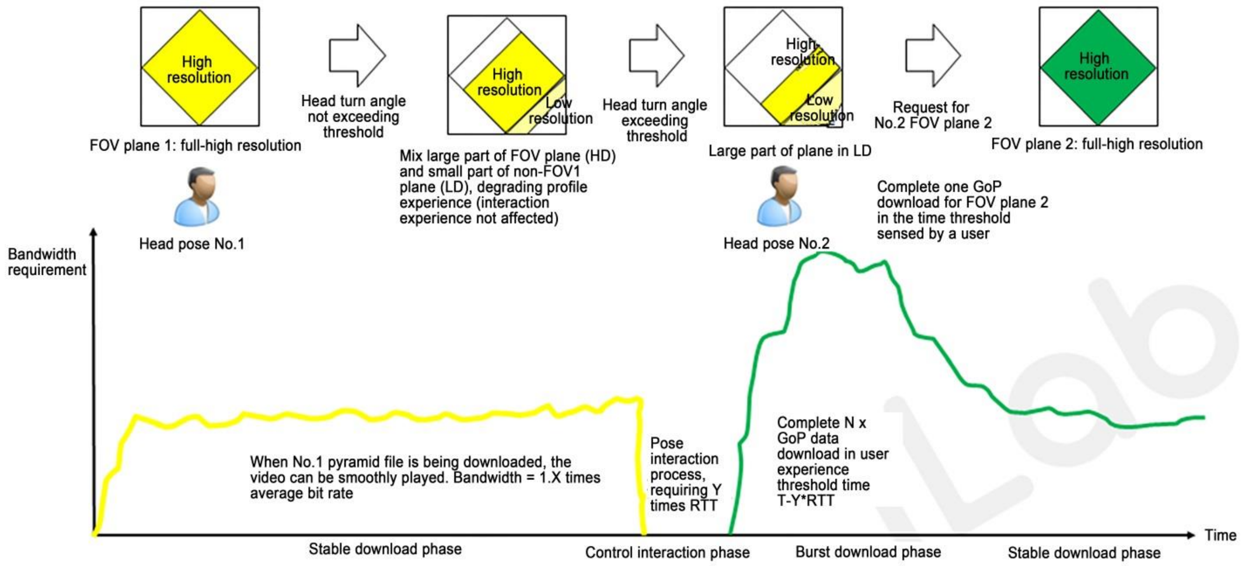
An FOV transmission scheme based on pyramid projection was published in Facebook in early 2016. Pyramid projection belongs to a kind of PSP projection technology, capable of reducing the average code-rate of media file to 20% of the original image quality of ERP projection. While it sacrifices some partial image-quality experience to reduce the requirement on the E2E 20ms interaction, as a modified FOV compromise scheme.



The above figure shows that all the spherical data corresponding to the user's visual information in virtual reality is put into the pyramid projection. The plane in right front of the user viewpoint is FOV plane, with high resolution code-rate; the rest four planes are non-FOV planes, with the resolution gradually lowering from side intersecting with FOV plane to the vertex in the opposite direction of visual angle. Adjust the pyramid after unfolded, and all the 360 °spherical visual information can be placed in the rectangle frame. This coding in rectangular frame format has very high compression efficiency; the code rate of pyramid projection can be reduced to 20% of the original image quality of ERP projection.

Regarding transmission technologies, Facebook uses the network transmission technology compatible with existing technologies to sacrifice some storage space for shortened latency and profile experience for interactive experience. This technology has following features:

1. It can number averagely distributed user head poses, and pre-create 30 tapered full-view files and save them on the server (sacrificing storage space for shortened latency), compatible with existing MPEG-DASH streaming solution.
2. When a user's head pose does not change, the high-resolution FOV plane is decoded by default.
3. When the change of a user's head position does not exceed the request threshold, the FOV information that the user receives is composed of the large part of FOV plane (high resolution) and small part of non-FOV plane (low resolution). In this way, this technology sacrifices profile experience to ensure the interactive experience.
4. When the change of a user's head position exceeds the request threshold, the FOV information that the user receives is composed of the small part of FOV plane (high resolution) and large part of non-FOV plane (low resolution). At the same time, the system requests the tapered full-view file mapping to the new pose and replaces the FOV information with FOV plane (high resolution) after Buffer obtains sufficient data, to sacrifice short-time profile experience for interactive experience.



3.5 Network Requirements Analysis

3.5.1 Online On-demand Service

The VR 360 online on-demand service is based on the Transmission Control Protocol (TCP) and its network requirements are determined by three factors: instant loading, smooth play, and view interaction. The view interaction for the full-view transmission solution is implemented at the terminal side, so that the network requirements are not considered.

1. Instant loading

   Instant loading is the time for a user to receive the virtual environment information when playing a on-demand VR 360 video. Instant loading is divided into three phases: instant loading signaling interworking (X1), minimum decoded buffer media packet download (Y), and player play loading preparation (Z). If a user's requirement for the instant loading time is **T1**, **X1+Y+Z≤T1**.

   If $\mathbf{Rate_{vr}}$ is set as the average bit rate of a VR 360 video, $\mathbf{Buffer_{time}}$ is the minimum decoded buffer media packet time of the terminal, $\mathbf{D_S}$ is the data volume at the TCP slow start phase, **X1*RTT** is the round-trip latency at the instant loading signaling interworking phase, **S*RTT** is the TCP slow start latency, and **Z** is the terminal player loading preparation latency, the end-to-end TCP throughput must meet the following formula:

$$\mathbf{TcpThrp_{min}} = \frac{Rate_{vr} * Buffer_{time} - Ds}{\mathrm{T1} - (\mathrm{X1} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

   If **P** is set as the packet loss rate (PLR), **BW** is the physical bandwidth, **MSS** is the minimum transmission unit, and **RTT** is the latency between the terminal and server, and based on the following classical TCP throughput formula:

$$\min(\mathrm{Max(BW)}, \frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}}) \geq \frac{Rate_{vr} * Buffer_{time}}{\mathrm{T1} - (\mathrm{X1} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

   we can obtain the following formula:

$$\min(\mathrm{Max(BW)}, \frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}}) \geq \frac{Rate_{vr} * Buffer_{time}}{\mathrm{T1} - (\mathrm{X1} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

2. Smooth play

   When a user watches a VR 360 video, the end-to-end TCP throughput should be always greater than N times the average bit rate, namely:

$$\min(\mathrm{Max(BW)}, \frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}}) \geq N * Rate_{vr}$$

   Where N is temporarily set to 1.5.

3. View interaction

   The view interaction requirement indicates the condition that the network must meet to make the Motion-to-Photons Latency (MTP) smaller than the target value when a user's view position changes, and minimize the user’s sense of FOV deterioration (black screen, information missing, or quality degradation). View interaction is divided into three phases: view change signaling interworking (X2), minimum decoded buffer media packet download (Y), and player play loading preparation (Z).

If a user's requirement on MTP is **T2** (often taking 20 ms) and the requirement for the time of FOV information deterioration during a head turn is **T3**, the equation **X2+Y+Z≤min(T2,T3)** must be met. This document analyzes Facebook's pyramid projection-based FOV transmission solution. Since the Facebook solution sacrifices the short-time profile experience for a better interaction experience, and the T2 requirement is also met on the terminal, we only need to meet the T3 requirement. If $\mathbf{Rate_{vr}}$ is set as the average bit rate of a VR 360 video, $\mathbf{Buffer_{time}}$ is the minimum decoded buffer media packet time of the terminal, $\mathbf{D_S}$ is the data volume at the TCP slow start phase, **X2*RTT** is the round-trip latency at the view change signaling interworking phase, **S*RTT** is the TCP slow start latency, and **Z** is the terminal player loading preparation latency, the end-to-end TCP throughput must meet the following formula:

$$\mathrm{TcpThrp_{min}} = \frac{Rate_{vr} * Buffer_{time} - Ds}{\mathrm{T3} - (\mathrm{X2} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

If **P** is set as the packet loss rate (PLR), **BW** is the physical bandwidth, **MSS** is the minimum transmission unit, and **RTT** is the latency between the terminal and server, and based on the following classical TCP throughput formula:

$$\min(\mathrm{Max(BW)}, \frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}}) \geq \frac{Rate_{vr} * Buffer_{time}}{\mathrm{T3} - (\mathrm{X2} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

we can obtain the following formula:

$$\min(\mathrm{Max(BW)}, \frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}}) \geq \frac{Rate_{vr} * Buffer_{time}}{\mathrm{T3} - (\mathrm{X2} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}$$

Based on above analysis, we summarize network requirements for the online on-demand service as follows:

	Bandwidth	Latency and PLR
Full-view transmission solution	$\geq \max(\frac{Rate_{vr} * Buffer_{time}}{\mathrm{T1} - (\mathrm{X1} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}, N * Rate_{vr})$	$\frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}} \geq \mathrm{Max(BW)}$
FOV transmission solution (Facebook)	$\geq \max(\frac{Rate_{vr} * Buffer_{time}}{\mathrm{T1} - (\mathrm{X1} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}}, N * Rate_{vr}, \frac{Rate_{vr} * Buffer_{time}}{\mathrm{T2} - (\mathrm{X2} + \mathrm{S}) * \mathrm{RTT} - \mathrm{Z}})$	$\frac{\mathrm{MSS}}{\mathrm{RTT}} \times \frac{1}{\sqrt{\mathrm{p}}} \geq \mathrm{Max(BW)}$

3.5.2 Online Live Telecasting

Network requirements for the VR 360 online live telecasting service are determined by three factors: instant loading, smooth play, and view interaction. The view interaction for the full-view transmission solution is implemented at the terminal side, so that the network requirements are not considered. This section describes the network requirements for UDP-based live telecasting of VR 360 videos. The live network requirements for TCP transmission are the same as the online on-demand service.

1. Instant loading

   Instant loading is the time for a user to receive the virtual environment information when joining a live channel of VR 360 videos. Instant loading is divided into three phases: instant loading signaling interworking (X1), a complete I-frame download phase (Y), and player play loading preparation (Z). If a user's requirement for the instant loading time is **T1**, **X1+Y1+Z1≤T1**.

Generally, the Fast Channel Change (FCC) solution is deployed to accelerate the live channel switching speed, and the bandwidth per user must not be less than 1.3 times the average bit rate, so that the FCC solution can function normally.

If $\mathbf{Rate_{vr}}$ is set as the average bit rate of the VR 360 video, **GopTime** is the Gop packet time, **T1** is the channel switching target value, $\mathbf{X1*RTT_{join}}$ is the round-trip latency at the signaling interworking (often for joining the multicast group) phase, and $\mathbf{T_{load}}$ is the terminal player loading preparation latency. the end-to-end UCP throughput must meet the following formula:

$$\mathbf{UdpThrp_{min}} = \mathbf{max}\left(\frac{Rate_{video} * GopTime * IFRatio}{\mathbf{T1} - \mathbf{X} * RTT_{join} - T_{Load}}, 1.3 * Rate_{vr}\right)$$

UDP-based live services are insensitive to the latency. However, the above formula shows that the denominator must be larger than 0, and **RTT** is usually not smaller than $\mathbf{RTT_{join}}$. Therefore, we have the following formula:

$$\mathbf{RTT} \geq RTT_{join} > \frac{\mathbf{T1} - T_{Load}}{\mathbf{X}}$$

2. Smooth play

No black screen or erratic display occurs when users are watching live VR 360 videos. Based on the TR-126 standard requirements, the network PLR is less than 10^{-6} and no erratic display occurs for live telecasting under the 4K resolution. Currently, peer vendors use the retransmission (RET) technology on the application layer to reduce video requirements on PLR. This lowers PLR to 10^{-4}. The network PLR requirements for higher resolutions are to be determined. That is:

$$PLR \leq \mathbf{PLR_{max}}$$

3. View interaction

The view interaction requirement indicates the condition that the network must meet to make the Motion-to-Photons Latency (MTP) smaller than the target value when a user's view position changes, and minimize the user’s sense of FOV deterioration (black screen, information missing, or quality degradation). View interaction is divided into three phases: view change signaling interworking (X2), minimum decoded buffer media packet download (Y), and player play loading preparation (Z).

If a user's requirement on MTP is **T2** (often taking 20 ms) and the requirement for the time of FOV information deterioration during a head turn is **T3**, the equation **X2+Y+Z≤min(T2,T3)** must be met. This document analyzes Facebook's pyramid projection-based FOV transmission solution. Since the Facebook solution sacrifices the short-time profile experience for a better interaction experience, and the T2 requirement is also met on the terminal, we only need to meet the T3 requirement. If $\mathbf{Rate_{vr}}$ is set as the average bit rate of the VR 360 video, **GopTime** is the Gop packet time, **T3** is FOV deterioration target value, **X2*RTTjoin** is the round-trip latency at the signaling interworking (often for joining the multicast group) phase, and $\mathbf{T_{load}}$ is the terminal player loading preparation latency. the end-to-end UCP throughput must meet the following formula:

$$\mathbf{UdpThrp_{min}} = \mathbf{max}\left(\frac{Rate_{video} * GopTime * IFRatio}{\mathbf{T3} - \mathbf{X} * RTT_{join} - T_{Load}}, 1.3 * Rate_{vr}\right)$$

We also have the following formula:

$$\mathbf{RTT} \geq RTT_{join} > \frac{\mathbf{T3} - T_{Load}}{\mathbf{X}}$$

Based on above analysis, we summarize network requirements for the live telecasting (UDP based) service as follows:

	Bandwidth	Latency and PLR
Full-view transmission solution	$\geq \max(\frac{Rate_{vr} * GopTime * IFRatio}{T1 - X * RTT_{join} - T_{Load}}, 1.3 * Rate_{vr})$	$RTT \geq RTT_{join} > \frac{T1 - T_{Load}}{X}$ $PLR \leq PLR_{max}$
FOV transmission solution (Facebook)	$\geq \max\left(\frac{Rate_{vr} * GopTime * IFRatio}{T1 - X * RTT_{join} - T_{Load}}, \frac{Rate_{vr} * GopTime * IFRatio}{T3 - X * RTT_{join} - T_{Load}}, 1.3 * Rate_{vr}\right)$	$RTT \geq RTT_{join} > \frac{T1 - T_{Load}}{X}$ $PLR \leq PLR_{max}$

3.6 Experience and Technology Evolution Route

The development of VR 360 video focuses on user experience and is a process of constant improvement of profile and information volume. Synchronization between the transmission technology and network technology determines the degree that the profile experience and interaction experience can reach.

We believe that the VR 360 video experience may evolve through following stages: early stage, entry-level experience stage, advanced experience stage, and extreme experience stage. Currently, VR still resides at the early stage, and has not reached the entry-level experience stage. In the following text, we have prejudged the terminal, content, experience, network, and arrival point for each stage.

3.6.1 Early Stage

VR 360 Videos for the early stage are called pre-VR. We believe that this stage should be marked by the highest hardware and software level that can be universally achieved in 2016, such as Samsung Gear for the HMD terminal and 4K VR 360 videos on Youtube.

1. Terminal:

   Angle of view: 90 °

   Screen resolution: 2K

2. Content:

   Full view resolution: 3840 x 1920

   Single-lens resolution: 960 x 960

   DoF: 2D (mostly)

   Color depth: 8 bit

   Frame rate: 30 fps

   Compression ratio: 165:1

   Coding standard: H.264

   Average bit rate: 16 Mbit/s

3. Experience:

   PPD: 11

Equivalent TV screen resolution: 240 P

Continuous experience time, as limited by the terminal and experience, does not exceed 20 minutes.

4. Network requirements for on-demand services:

   Bandwidth for smooth play: **25 Mbit/s** (based on the full-view transmission solution)

5. Network requirements for live telecasting:

   Bandwidth for smooth play: 20.8 Mbit/s (based on the full-view transmission solution)

3.6.2 Entry-Level Experience Stage

VR 360 Videos for the entry-level stage are called entry-level VR. We believe that this stage will last about two years. The highest hardware and software level that can be universally achieved in 2016 is further promoted: The terminal screen resolution is improved to 4K and the full-view resolution is improved to 8K. As a result, the profile that users receive approaches the 480P PPD effect on PC.

1. Terminal:

   Angle of view: 90 °

   Screen resolution: 4K

2. Content:

   Full view resolution: 7680 x 3840

   Single-lens resolution: 1920 x 1920

   DoF: 2D (mostly)

   Color depth: 8 bit

   Frame rate: 30 fps

   Compression ratio: 165:1

   Coding standard: H.264

   Average bit rate: 64 Mbit/s

3. Experience:

   PPD: 21

   Equivalent TV screen resolution: 480 P

   Continuous experience time, as limited by the terminal and experience, does not exceed 20 minutes.

4. Network requirements for on-demand services:

   Bandwidth for smooth play: **100 Mbit/s** (based on the full-view transmission solution)

5. Network requirements for live telecasting:

   Bandwidth for smooth play: 83.2 Mbit/s (based on the full-view transmission solution)

3.6.3 Advanced Experience Stage

The VR 360 video of advanced experience stage is called Advanced VR. At this stage, the screen resolution, chip performance, and ergonomics of HMD terminal and the quality of content are largely improved. We expect the achievement of this objective after 3 to 5 years' industrial development.

1. Terminal:

   Angle of view: 120 °

Screen resolution: 8K

2. Content:

   Full view resolution: 11520*5760

   Single-lens resolution: 3840*3840

   DoF: 2D (mostly)

   Color depth: 10 bits (HDR)

   Frame rate: 60 fps

   Compression ratio: 215:1

   Coding standard: HEVC/VP9

   Average bit rate: 279 Mbit/s

3. Experience:

   PPD: 32

   Equivalent TV screen resolution: 720 P

   Maximum continuous experience duration: 20–60 minutes

4. On-demand network requirements:

   Bandwidth requirement for smooth play: **418 Mbit/s** (based on full view transmission scheme)

5. Live broadcast network requirements:

   Bandwidth requirement for smooth play: 361.4 Mbit/s (based on full view transmission scheme)

3.6.4 Ultimate Experience Stage

The VR 360 video of ultimate experience stage is called Ultimate VR. At this stage, the development of HMD terminal and the content can provide users with best usage experience. We expect full access to this stage after 10 years' long-term industrial development.

1. Terminal:

   Angle of view: 120 °

   Screen resolution: 16K

2. Content:

   Full view resolution: 23040*11520

   Single-lens resolution: 7680*7680

   DoF: 3D (mostly)

   Color depth: 12 bits

   Frame rate: 120 fps

   Compression ratio: 350:1

   Coding standard: H.266

   Average bit rate: 3.29 Gbit/s

3. Experience:

   PPD: 64

   Equivalent TV screen resolution: 4K

   Maximum continuous experience duration: over 60 minutes

4. On-demand network requirements:

Bandwidth requirement for smooth play: 4.93 Gbit/s (based on full view transmission scheme)

987 Mbit/s (based on Facebook scheme, and It will be the mainstream in this stage)

Instant interactive bandwidth requirement: 2.35 Gbit/s (based on Facebook scheme)

5. Live broadcast network requirements:

Bandwidth requirement for smooth play: 4.277 Gbit/s (based on full view transmission scheme)

856 Mbit/s (based on Facebook scheme)

Instant interactive bandwidth requirement: 2.35 Gbit/s (based on Facebook scheme)

3.6.5 Summary of Evolution Route

At the entry-level experience stage, the entry-level VR profile experience is not very good; therefore, the industry will choose the full view transmission scheme to ensure good interactive experience. At this stage, the penetration rate and absolute number of VR users will not be very large. From the view of user's end-to-end network bandwidth requirement, the bandwidth good enough for smooth play of 8K video can also meet the requirement of VR 360 video.

At the advanced experience stage, the Advanced VR profile experience has been enhanced; therefore, the full view transmission scheme has more strict requirement on network bandwidth, and it can still ensure good interactive experience as long as the network bandwidth is ready.

For Ultimate VR, the single-lens profile has reached the retina level, and the requirement on network bandwidth is too strict for the full view transmission scheme. Therefore, we have to adopt FOV transmission, which has certain requirement on low latency of the network. If the compromised FOV scheme of Facebook is adopted, part of the profile experience will be compromised and the requirement on network bandwidth and latency will be lowered.

At the entry-level experience stage and the advanced experience stage, the continuous experience duration acceptable to users is not long. Compared with other traditional business, the huge bandwidth demands of a family user is short. While at the ultimate experience stage, the sudden bandwidth requirement triggered by the FOV transmission scheme during head turning is much more strict than that during stable watch without head turning. Therefore, at all stages of the evolution of VR 360 video, there is always short-time strict requirement on bandwidth based on the demands.

In summary, our judgment on the evolution route of VR 360 video can be summarized as follows:

Standard	Pre-VR	Entry-Level VR	Advanced VR	Ultimate VR
Continuous experience duration	Less than 20 minutes	Less than 20 minutes	20 to 60 minutes	Over 60 minutes
Estimated time	Now – 2 years	Now – 2 years	3–5 years	5–10 years
Video resolution	Full view 4K 2D video (Youtube) (Full frame resolution 3840*1920)	Full view 8K 2D video (Full frame resolution 7680*3840)	Full view 12K 2D video Full frame resolution 11520*5760	Full view 24K 3D video Full frame resolution 23040*11520

Standard	Pre-VR	Entry-Level VR	Advanced VR	Ultimate VR
Single-lens resolution	960*960 [with glasses and view angle of 90 °]	1920*1920 [with glasses and view angle of 90 °]	3840*3840 [with head-mounted display and view angle of 120 °]	7680*7680 [with head-mounted display and view angle of 120 °]
PPD (Note 1)	11	21	32	64
Equivalent traditional TV screen resolution	240P	480P	720P	4K
Color depth (bit)	8	8	10 (HDR)	12
Compression ratio (Note 2)	165:1	165:1	215:1 (Note 2)	350:1 (3D) (Note 2)
Frame rate	30	30	60	120
Typical video bit rate	16 Mbit/s	64 Mbit/s	279 Mbit/s	3.29 Gbit/s
Typical network bandwidth requirement (Note 3)	**25 Mbit/s**	**100 Mbit/s**	**418 Mbit/s**	**1Gbit/s** *for smooth play* **2.35Gbit/s** *for instant interactive* (Note5)
Typical network RTT requirement (Note 4)	**40 ms**	**30 ms**	**20 ms**	**10 ms**
Typical network packet loss requirement (Note 4)	**1.4E-4**	**1.5E-5**	**1.9E-6**	**5.5E-8**

Note 1: PPD is short for Pix per Degree. The retina of common human beings can reach the resolution of 60 PPD.

Note 2: The compression is calculated based on the empirical value of H.264, HEVC, and H.266 development. The contents of left and right 3D eyes are highly related; therefore, large compression is okay and will not comprise the quality too much.

Note 3: Typical network bandwidth. The on demand is estimated based on 1.5 times the bit rate.

Note 4: Network latency and packet loss. Determine the target latency value first and then calculate the packet loss based on the network bandwidth and the TCP throughput formula.

Note 5:In this stage FOV transmission will be the mainstream.

4 VR Service-oriented Bearer Network

4.1 Analysis on Network Impacts from VR Service

4.1.1 Traffic Characteristics of VR

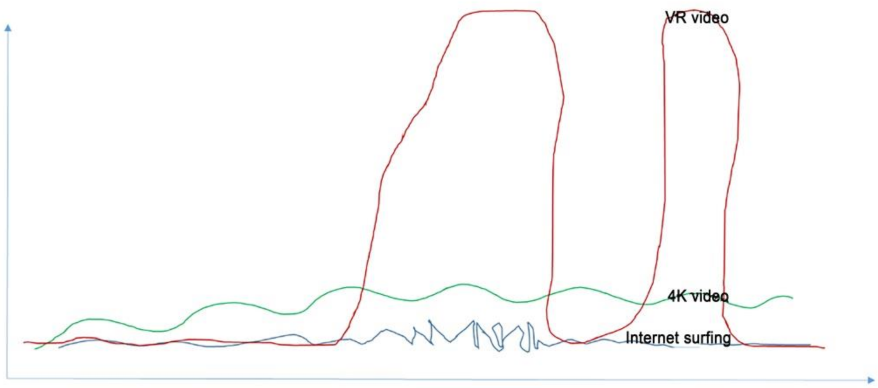
If the broadband Internet access is like a stream with frequent seasonal cutoff, the video (including HD and 4K/8K video) is like the Eerqisi River and VR is like the Yangtze River or Amazon River. Their demands of different services for pipe bandwidth vary greatly from ten times to a hundred times. For common broadband access, great surfing experience happens when the (temporary) peak bandwidth reaches 20 Mbit/s to 30 Mbit/s. However, for HD and 4K/8K video, good surfing experience must be supported by continuous 30 Mbit/s to 100 Mbit/s bandwidth, and for VR video, excellent experience must be supported by home bandwidth faster than one Gbit/s.

However, as the saying goes, Rome was not built in a day. To obtain ultimate VR experience, the whole industry chain must experience a slow and continuous progressing process, either in terminals, network, contents, or development of potential users. Considering VR is an important service in the future society, as the pipe for delivering contents, it is required to analyze the differences between VR services and current pipes in advance, to ensure that the bearer network gets ready before VR service is launched and lay a good foundation for VR service. Detailed analysis on features of VR service steam are as follows.

Characteristic 1: VR video is typical elephant flow.

Compared to the existing Internet access and video services, VR is absolutely the legendary elephant flow, which penetrates the whole metropolitan area network and access network (in the past, the elephant flow probably exists in DCI, metropolitan area exit, and backbone network only). Transmission of one channel of VR contents requires multiple end-to-end Gbit/s bandwidths and such bandwidths will extend to the user home. Meantime, as the VR video is watched by individuals independently (the VR glasses is the most common terminal), the possibility that multiple channels of VR services exist in the same user home at the same time increases greatly. Compared to the family watching of 4K/8K video through large-size screen, the bandwidths consumed by VR service may double continuously. Though the transmission bandwidth for one channel of VR service may be further reduced due to future technological development and improved compression algorithm, there is a great possibility that the home bandwidth is faster than one Gbit or even 10Gbit/s. Compared to the bandwidths consumed by VR video, that consumed by other services can be ignored basically.

Figure 4-1 Comparison of bandwidths consumed by major network services



Characteristic 2: more remarkable tidal effect in network traffic

Restricted by the business mode and VR device (long-time wearing may cause dizziness, syncope, and other discomfort), the continuous online time of users may not be too long, which along with the abovementioned elephant flow certainly will cause more huge difference between traffic crest and valley, especially in case of hot events (100 times difference). For the BRAS/CR devices integrated in the whole metropolitan area and the links carried the traffic, dynamic application for link resources and flexible expansion or shrinkage of device capacity as required are very important features.

Characteristic 3: higher requirements for latency and packet loss rate

The VR service has high requirements for bandwidth, network latency, and packet loss rate. In terms of transmission of VR video over network, except for the live streaming, most videos on demand are transmitted based on TCP (OTT VR live streaming is also transmitted based on TCP). Latency and packet loss impacting the TCP throughput are even more important in transmission of VR service of large code stream. The network is liable for reduction of latency and packet loss rate.

For the latency, from the end-to-end perspective, except for the transmission latency of light in optical fiber caused by distance (The latency caused by light in optical fiber transmission is hard to be eliminated in short time. The optical speed in fiber transmission is 2/3 of that in vacuum transmission, equaling to about 200 KM/ms), the forward latency caused by devices during forwarding has a greater impact. During the forwarding process, the router works as a storage and forwarding mechanism, which will cause great queuing latency when traffic bursting causes queuing in inbound and outbound directions. However, in the statistical multiplexing IP network, congestion and queuing are also inevitable. Such congestion includes the congestion induced by that the actual service traffic exceeds the actual physical bandwidth limit, and the instantaneous congestion (false congestion) caused by the service itself due to burst mechanism. Both types of congestion may cause great latency and even packet loss in serious cases.

The packet loss has an even greater impact on TCP throughput. Of course, the aging of links (optical power attenuation and dust in and wear of optical plugs) and storage jumping of device also may cause packet loss, which are rare events after all. Compared to the optical fiber transmission, retransmission often occurs in copper wire and Wi-Fi transmission as they have to solve the unreliable transmission problem (interference caused packet loss). This seriously restricts the TCP throughput and causes greater latency.

4.1.2 Estimation on bandwidth of families using VR service

Entry-level experience stage (Entry-level VR):

VR video: bandwidth of single program: 100 Mbit/s

Ordinary video: two 4K video programs, 25 Mbit/s for each program.

Internet access service: 10 Mbit/s bandwidth

Total bandwidth of family: 160 Mbit/s. The home bandwidth of family should be over 200 Mbit/s.

Advanced experience stage (advanced VR):

VR video: bandwidth of single program: 400 Mbit/s; multi-screen may be used (two screens are considered).

Ordinary video: two 4K video programs, 25 Mbit/s for each program.

Internet access service: 20 Mbit/s bandwidth.

Total bandwidth of family: 870 Mbit/s. The home bandwidth of family should be 1 Gbit.

Ultimate experience stage (ultimate VR):

VR video: bandwidth of single program: 2.35 Gbit/s, calculated as 2 screens.

Other services: ordinary video and Internet access services: 200 Mbit/s

Total bandwidth of family: 4.9 Gbit/s. The home bandwidth of family should be 5 Gbit/s (this is calculated based on FOV mode and this applies to the following scenarios).

4.1.3 Network Impacts of VR Service

As an emerging service, VR service certainly will have impacts on the current network architecture as it has great demands for network transmission bandwidth. This section mainly describes the impacts of VR service on such aspects as Internet access bandwidth, access network, metropolitan area network, and home network.

The impacts on network mainly cover such three aspects as **devices of higher capacity**, **new service transmission program**, and **new network architecture**.

1. Home bandwidth: Gbit/s, faster than one Gbit/s, 10 Gbit/s (ultimate VR and multi-channel VR)
2. Access network: 10G/40G/100G PON large-area deployment. A single upstream OLT link supports 400G ports.
3. Metro network: 10TB-level board, PB-level device switching capacity.
4. Devices and links support dynamic capacity expansion: 100 times difference between crest and valley of tide. Application for devices and link resources as required can significantly reduce the operation cost (power consumption and noise).
5. Home network: It has much higher requirements for Wi-Fi, such as 802.11ad, snap high-speed transmission, and support of multicast or even other wireless transmission protocols by Wi-Fi.

4.2 Home Network

VR's watching devices determine that VR service is mainly used at home. As the high requirement for bandwidth and more and more terminals owned by a family, it is possible that different users stay in different rooms. Therefore, the home network mostly based on Wi-Fi has extremely high requirements for bandwidth, latency, and coverage. The current home Wi-Fi still faces great challenges.

Problems existed in home Wi-Fi bearer network of VR service

- Super high bandwidth cannot be satisfied: Most Wi-Fi networks used by current families are 802.11n and the maximum bandwidth is about 300 Mbit/s. However, the bandwidth required for each Advanced VR video is over 400 Mbit/s and that for future Ultimate VR video is 2.35 Gbit/s (FOV mode) or even 5 Gbit/s (full view mode).
- Mutual impacts of different services: Except for the VR video service, the Internet access, downloading, and other non-VR video services over the same network will preempt resources and impact each other;
- Signal coverage: Objects inside the house will attenuate the Wi-Fi signal, especially the load bearing walls and metals, which will cause different signal intensity in different locations of a house and even no signal in certain locations;
- Channel congestion and interference: In urban areas, the 2.4 Gbit/s frequency band widely used in residential buildings has too much signal sources and such signals will impact each other. Meantime, such wireless devices as cordless telephones, microwave ovens, and Bluetooth will have impacts on the wireless signals of 2.4 Gbit/s frequency band.

Solutions to home network with ultra high bandwidth, high performance and full Wi-Fi coverage

- New Wi-Fi standard equipment: Advanced VR users need commercial 802.11ac Wi-Fi routers which feature an air interface bandwidth ranging between 1 Gbit/s and 1.3 Gbit/s. Ultimate VR users need an access to 5 Gbit/s bandwidth in their houses, so they need 802.11ac Wave2 or 802.11ad or 802.11ax routers. The 802.11ac Wave2 routers have a maximum air interface bandwidth of 6.93 Gbit/s and a faster transmission rate. The 802.11ad routers (also called WiGig routers) have a maximum throughput of 7Gbit/s. Another option is the routers supporting 802.11ay Wi-Fi standard which is still being discussed.
- Differentiate bearer in accordance with business: Select proper frequency band and SSID in accordance with service feature. For example, 2.4 Gbit/s Wi-Fi is possible for general internet business with low requirement on internet; 5 Gbit/s Wi-Fi is possible for normal video, 4K and Advanced VR with relatively high requirement on internet; 60 Gbit/s Wi-Fi(802.11ad/802.11ay) or 5 Gbit/s Wi-Fi(802.11ac Wave2/802.11ax) is possible for Ultimate VR.
- Wi-Fi distributed plan: Wireless network has poor capacity in object penetration in high frequency band. Arrangement of home Wi-Fi internet is converted from centralized mode to distributed mode and several Wi-Fi access points (AP) is deployed, to reach ideal connection speed and to meet different demand of terminals at any place of home. Three distributed access modes, including Wi-Fi Repeater, PLC AP and ETH AP, allow flexible expansion and plug and play, and supports the experience of seamless switching. The three modes are possible for Advanced VR only if bandwidth requirement is met. ETH AP is recommended for Ultimate VR service.
- Minimization of interference with wireless signal: It mainly includes two aspects. First, selected channel shall be different from that used by other AP in current circumstance, to reduce interference. Second, household appliances, especially electric appliance

vulnerable to interference (such as microwave oven) shall keep a certain distance from Wi-Fi source. Meanwhile, electric appliance with no EMI certification is forbidden in home use.

- Support of multicast in a better way: Support DMS (Directed Multicast Service) and Groupcast with Retries (GCR), and transmit multicast data better in Wi-Fi condition.

Summary: requirement of VR service on home Wi-Fi network

- **Next generation of Wi-Fi protocol standard**, broader frequency spectrum and wide application of MU-MIMO and Beamforming technologies: 802.11ac/802.11ad/802.11ax/802.11ay, 5G/60G frequency band;
- Support live service in a better way, Wi-Fi required to better support multicast, and terminal required to provide multicast function
- **Supply of service to differentiate capacity of bearer**: Different service employs different SSID and frequency band for isolation
- **Distributed Wi-Fi plan improves signal coverage in the whole family**

4.3 RAN

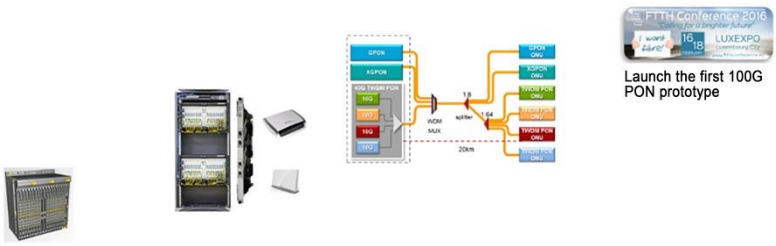
In the whole network, access network, as the operator's nearest network to users, collects all end users. Following continuous development of technologies, now access technology mainly includes copper wire access to FTTB / C and fiber access to FTTH. Copper wire access is divided into ADSL, VDSL2, Vectoring, Super Vectoring and G.fast, and fiber access divided into GPON and EPON access.

The following analyzes access technologies for VR video service.

4.3.1 Analysis on Fiber Access Technology

Fiber access technology plays an important role following gradually increasing requirement of human on bandwidth. From original GPON with uplink of 1G and downlink of 2.5G and EPON with uplink of 1G and downlink of 1G, now GPON with uplink of 2.5G and downlink of 10G and EPON with both uplink and downlink of 10G have been used for commercial purpose in certain scale. 40G PON is mature, and is also to be used for commercial purpose until cost problem is solved. Meanwhile, 100G PON also has mockup. Ultra high bandwidth and long transmission distance make fiber access become the first option in new construction and transformation of access network. The following describes technologies used for VR video service in FTTH.

Figure 4-2 Evolution of Fiber Access Technology



G PON: In basic experience stage of VR, G PON is mainly for ordinary video service. Household inbound bandwidth and downlink bandwidth of port are respectively 200M and 1G.

Each port involves with 16 users as calculated by taking coincidence rate as 30%. It is recommended that splitting ratio is set up to 1:16. When household inbound bandwidth is 1 Gbit/s in advanced experience stage of VR, GPON port is not recommended as it has no advantage.

10G PON: In basic experience stage of VR, 10G PON is sufficient for 200M household inbound bandwidth. In advanced experience stage of VR, household inbound bandwidth is 1 Gbit/s. 10G port involves with 33 users as calculated by taking coincidence rate as 30%. It is recommended that splitting ratio is set as 1:32. In ultimate experience stage of VR, household inbound bandwidth is up to 5G. At this time, extremely low splitting radio of 10G PON would result in cost rise. So it is recommended to smoothly progress to 40G/100G PON.

40G PON: It is not considered in basic experience stage of VR as it is not used for commercial purpose now, but it is assumed that 40G PON is to be used for mature commercial purpose in advanced experience stage of VR. In advanced experience stage of VR, inbound bandwidth of a single user is 1G. Each port involves with 100 users as calculated by taking coincidence rate of 40%. Splitting ratio is configured as 1:64 or 1:128 in accordance with actual conditions. In ultimate experience stage of VR, bandwidth of a single user is up to 5G. Each port involves with 20 users as calculated by taking coincidence rate of 40%. It is recommended to set splitting ratio as 1:16.

100G PON: Now 100G PON only has mockup and has a long way to be used for commercial purpose in certain scale, with relevant standard under preparation. It makes progress in combination with VR development process. Only network design where 100G PON supports ultimate VR in the future is analyzed. Household inbound bandwidth for ultimate VR is 5G in FOV mode. Each port involves with 50 users as calculated by taking coincidence rate of 40%. It is recommended to set splitting ratio up to 1:64. However, bandwidth may be obtained in FOV mode by giving up experience in head rotation moment. Therefore, In case that bandwidth is high enough in the future, direct transmission of full-view video could gain better experience. In this circumstance, single connection bandwidth is 4.93G and household inbound bandwidth is 10G. Each port involves with 25 users as calculated by taking coincidence rate of 40%. Splitting ratio is set as 1:16.

OLT framework: Following video popularization, bandwidth required by users and coincidence rate greatly increase. OLT, as core node to network, faces great challenge in handling capacity. Traditional concentrated framework, which carries out centralized forwarding via master-control switching chip, has become the bottleneck of the whole equipment. Distributed framework is required to distribute service to service board. Meanwhile, reliability is improved. In addition, to handle bursting of video streaming, interface board shall also has certain caching capacity (ONT is also required to increase cache) to prevent blocking from packet loss.

4.3.2 Analysis on Copper Access Technology

ADSL/VDSL2: Downlink speed of ADSL and VDSL2 is respectively up to 25 Mbit/s and 50 Mbit/s. The speed is possible for surfing the internet, but for video service, 30 Mbit/s to100 Mbit/s required by 4K/8K video is impossible, let alone VR video in Gbit/s. Therefore, it could not meet the requirement.

Vectoring: On the basis of VDSL line, downlink bandwidth is up to 120 Mbit/s through improvement of line speed by offset of far-end crosstalk. Although 4K/8K video could be seen, household bandwidth is 200 Mbit/s in basic stage with VR video included. Therefore, it is not recommended.

SuperVectoring: Expand working frequency band from 17MHz to 35MHz based on Vectoring. This increases downlink bandwidth, which is 300 Mbit/s within 300m and 100 Mbit/s in 300–700m. SuperVectoring could be used in basic experience stage of VR, but it fails to meet bandwidth requirement in advanced and ultimate experience stage following further increase of resolution ratio.

G.fast: Expand working frequency band to 106MHz based on Vectoring technology, to make transmission bandwidth up to 1Gbit/s, but ensure that bandwidth is transmitted up to 100m. This meets the deployment from fiber to port. However, as bandwidth of G.fast 1Gbit/s is equal to uplink bandwidth plus downlink bandwidth, actual downlink bandwidth is less than 1 Gbit/s in actual conditions (such as, uplink of 200 Mbit/s, downlink of 800 Mbit/s, adjustable). When surfing the internet and playing 4K/8K video service are carried out at the same time, an Advanced VR video could be seen at best.

The above analysis result indicates that copper wire access technology, besides SuperVectoring and G.fast, is not suggested to be deployed in the network where VR video is only played. SuperVectoring is only used in basic experience stage of VR and G.fast is proper for Advanced VR, but synchronous multiplex playing will still result in insufficient bandwidth. Although copper wire could reduce cost, PON access line is to be deployed for long-term purpose, to meet video application which has increasing requirement in the future.

In addition, working frequency band is increased following bandwidth in employment of copper wire access technology, but frequency increase will result in acute reduction of efficient transmission distance. Meanwhile, support scope of each access equipment greatly decreases and long latency is found as compared with fiber access technology. The above indicates that unless new breakthrough is made. copper wire access technology is not proper in future VR service network which has high requirement on bandwidth.

4.3.3 Summary: Requirement of VR Video on the Access Network

Table 4-1 VR support on access technologies

	Split ratio	Downlink bandwidth	Basic VR (100 Mbit/s)	Advanced VR (418 Mbit/s)	Ultimate VR	
					2.35 Gbit/s (FOV)	4.93G(full view)
GPON	1:32	30 Mbit/s				
	1:64	10 Mbit/s				
10G PON	1:32	300 Mbit/s	√			
	1:64	100 Mbit/s	√			
40G PON	1:32	1 Gbit/s	√	√		
	1:64	500 Mbit/s	√	√		
100G PON	1:16	6 Gbit/s	√	√	√	√
	1:32	3 Gbit/s	√	√	√	
	1:64	1.5 Gbit/s	√	√		
G.fast	/	800 Mbit/s	√	√		
Super Vectoring	/	300 Mbit/s	√			
Vectoring	/	120	√			

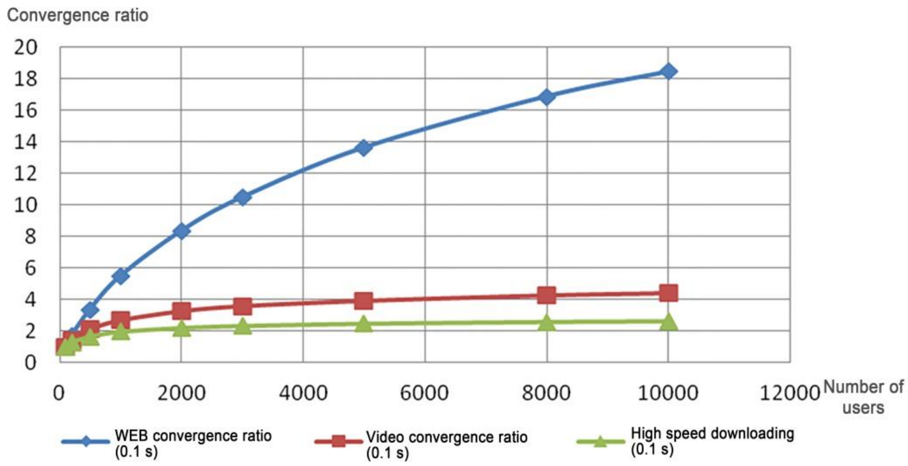
Note: 1 Gbit G.fast bandwidth is equal to uplink bandwidth plus downlink bandwidth, which is respectively 200 Mbit/s and 800 Mbit/s.

- In basic experience stage of VR, household inbound bandwidth is 200 Mbit/s. Therefore, both SuperVectoring, G.fast and 10G PON access technology could be adopted. As distance and cross talk has great influence on copper wire, service shall not be opened until evaluation on line quality is conducted. Inbound bandwidth is at least 100 Mbit/s.
- In advanced experience stage of VR, household inbound bandwidth is 1 Gbit/s. Therefore, 40G PON technology could be adopted.
- In ultimate experience stage of VR, household inbound bandwidth is 5 Gbit/s. Therefore, future 100G PON access technology is required.
- Following equipment's increasing handling capacity, centralized OLT framework shall be converted to distributed type, to break through bottleneck. The interface board has certain caching capacity.

4.4 Metropolitan Area Network (MAT)

Following development of VR and 4K/8K video services, user's bandwidth can reach or exceed Gbit/s. Traffic of MAT may surge by ten, hundred and even thousand times. Meanwhile, long durability of video service results in drop of uplink and downlink convergence ratio in the whole network.

Figure 4-3 Comparison of Convergence Ratio of Different Services



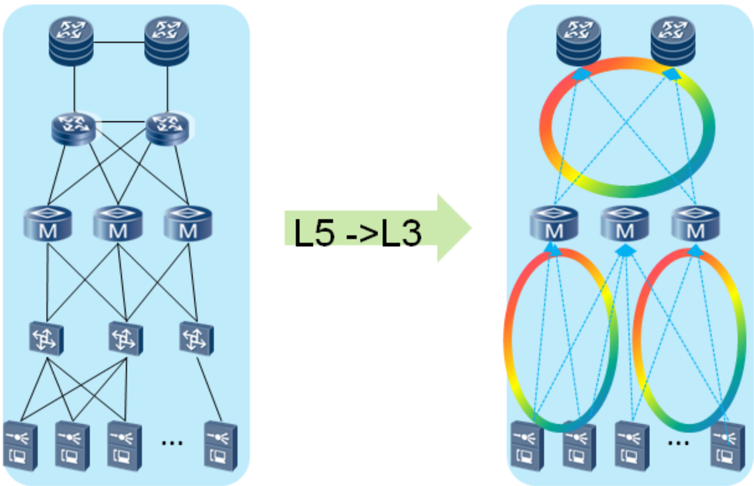
Acute traffic increase, change in convergence model and high requirement of VR and 4K videos on bandwidth, latency and packet loss challenges bearer of traditional network and high convergence.

- **Low network efficiency**: Expansion of port-to-port network equipment is needed after video traffic increases. Many convergence layers make low convergence ratio, and port-to-port equipment with synchronous expansion increases in terms of scale. CDN is deployed in high position and away from the end user, and service stream travels through much network equipment. Therefore, it occupies plenty of network resources. In addition, more equipment number increases the probability of bottleneck and blocking and port-to-port latency.

- **Poor user experience:** In case that services are provided at the same time, packet loss and latency synchronously rise following increase of network utilization, reducing video service experience. In low load network, 98.7% sudden packet loss occurs at transitional convergence node from high bandwidth to low bandwidth. Increase of packet loss reduces video service experience.

Based on reasons mentioned above, first simplify layer and structure of traditional network and move CDN down to BNG or even CO, to improve transmission efficiency of the whole MAT.

Figure 4-4 Simplification of Framework of MAT



Main characteristics of the framework: Eliminate convergence layer of LSW and MAT; BNG directly connects up to CR and drops down to network edge; and OLT directly connects to BNG and OTN to CO.

As network is simplified, capacity of equipment shall be further improved. The following evaluates model of simplified MAT:

- Uplink and downlink convergence ratio of OLT: 1:2
- Uplink and downlink convergence ratio of BNG: 1:2
- Uplink and downlink convergence ratio of CR: 1:2
- Each OLT involves with 4000 users
- Each BNG involves with 20 OLTs
- Each CR involves with 10 BNG

Table 4-2 Evaluation on Traffic Model

		Entry-level VR	Advanced VR	Ultimate VR
Requirement on home bandwidth		**200**	**1000**	**5000**
Penetration rate		15%	30%	40%
Coincidence rate		10%	20%	30%
Bandwidth	**CR uplink (G)**	150	6000	6000

		Entry-level VR	Advanced VR	Ultimate VR
requirements	**CR downlink (G)**	300	12000	120000
	BNG uplink (G)	60	1200	12000
	BNG downlink (G)	120	2400	24000
	OLT uplink (G)	6	120	1200
	OLT downlink (G)	12	240	2400
	CR(G)	450	7020	180000
	BNG(G)	180	3600	36000
	OLT(G)	18	360	3600

As shown in the above table, as video service is popular, there is high requirement on handling capacity of equipment with MAN adopted.

BNG: In advanced experience stage of VR, integrated machine's handling capacity is required to be 3.6T; in ultimate experience stage of VR, integrated machine's handling capacity is required to be 36T. If chain utilization rate of not more than 70% is considered, integrated machine’s handling capacity is 5T and 50T respectively in the above two stages.

CR: In advanced experience stage of VR, integrated machine's handling capacity is required to be 7T; in ultimate experience stage of VR, integrated machine's handling capacity is required to be 180T. Actual handling capacity of CR is 10T and 257T respectively in the two stages as calculated by taking chain utilization rate of 70%.

The above CR capacity is calculated out considering each CR involves with 10 BNG, but now only two CRs (backup and sharing for each other) connect to backbone network in many cities. According to the network model, in a city with a population of 5 million, CR capacity will reach 5000000/4000/20*180T/10=1125T in ultimate experience stage of VR. Considering chain utilization rate is kept under 70% and CR capacity is more than 1607T to prevent from blocking and packet loss during actual use, P bit level super router is required.

Although network capacity continuously improves in terms of equipment and chain following increasing requirement of VR on network and if VR video with high bandwidth and low latency and packet loss is supported on the network with no difference, ultimate experience shall not be guaranteed due to traffic burst and resource occupation by different services. Considering characteristics of VR services, the following specifically describes some technologies to further improve user experience.

4.5 Network Solution of Supporting FOV Transmission Technology

According to the difference in real-time display and transmission mode, video transmission can be divided into unicast-based on demand and multicast-based live broadcast. The analysis of VR FOV transmission based on the above mentioned two modes is as follow.

4.5.1 VR On-demand Network Solution

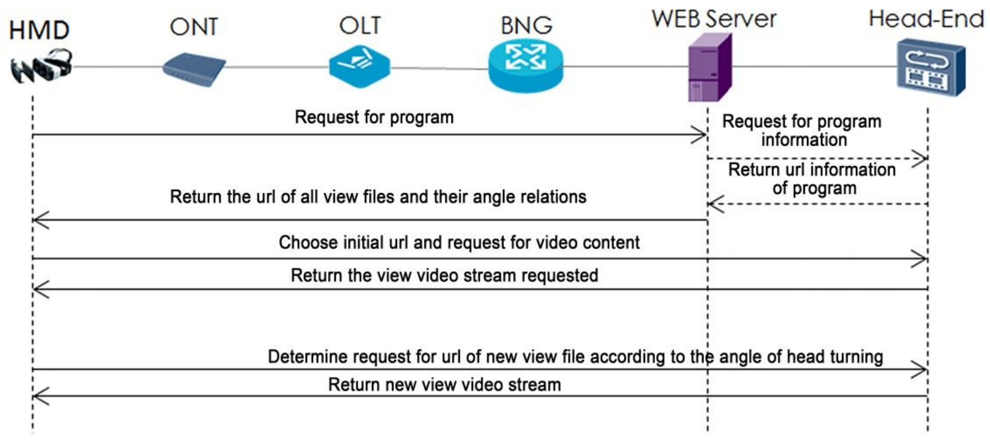
Since FOV transmission mode has several view files, server needs to maintain the URL of each view file and angle ration information. When client request for program initially, server will send such information to the client. Client can use the initial view file suggested by server as the initial viewing video stream or chose on its own.

Table 4-3 Views and URLs of VR on-demand (taking 12 views as an example)

Program	URL	Angle	Whether Initial View
Program a	URL a1	0	**Yes**
	URL a2	+30	
	URL a3	+60	
	……	……	
	URL a12	+330	
Program b	URL b1	0	**Yes**
	URL b2	+30	
	URL b3	+60	
	……	……	
	URL b12	+330	

Once initial view file is chosen, current view shall be used as reference to maintain view files of other views. Corresponding view files can be calculated accurately when turning head to a new view.

Figure 4-5 FOV VR on-demand process



Program request: Clint will send request to server to display videos according to the link on the access page. In practice, this interaction is usually in the process that client send request to

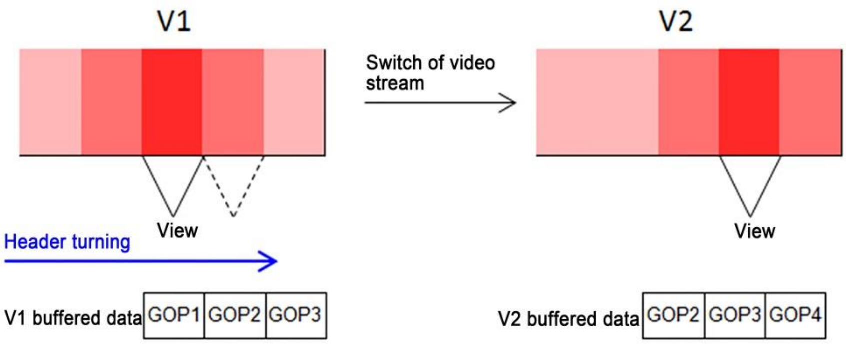
WEB server to display video and WEB server return the URL of the video to client, and then client access server to download video stream according to the acquired URL.

For videos transmitted by FOV VR, there will be several URLs. What returned by WEB server is the URLs of all view files and angle relationship information contained in the requested program package.

Head turning: When receiving the initial URL requested, client shall record the current view and map such URL to local views according to the angle relations between each view file acquired from WEB server. If head turned during the viewing process, client will request to video server for video files of new view using corresponding URL and time offset information.

When view changes due to head turning, client shall request for video stream of new view. As shown in the following figure, V1 is the video stream of original view and V2 is the video stream of new view. Before the arrival of V2 video stream, client will continue to display the buffered content of V1 to avoid interruption of display. When the buffered data of V2 video meets the display conditions, client will change to display the content of V2.

Figure 4-6 Switching of view video stream when turning head during on-demand



The handling procedures from head turning to display of buffered data of new view includes the following parts:

A: Perception by sensor and handling

B: Switching from request for data of V1 to request for data of V2

C: Download data of V2

D: Load and display data of V2

The requesting initial position of V2 video stream varies according to the difference in current viewing position of V1:

1. If the GOP of current V1 just starts to display and the remaining time is enough to download the minimum amount of buffered data for display, client will start to request data of V2 from the next GOP of current displayed GOP of V1.
2. If the remaining time of displaying the current GOP of V1 is limited and not enough to download the minimum amount of buffered data for display, client will start to request data of V2 from the GOP after the next GOP of current played GOP of V1. For example, if GOP1 of V1 is currently displayed and the minimum amount of data of V2 for display cannot be downloaded before finishing displaying of GOP2, client will request for data

of V2 from GOP3 and wait to display GOP3 of V2 after finishing displaying of GOP2 of V1.

For the above described assumption, the minimum unit of switching displaying from V1 to V2 is GOP. If player can make frame-level switch, then after finishing download the minimum amount of buffered data of V2 and before finishing displaying of current GOP of V1, client may switch the display content to V2 for better experience. However, frame-level switch requires more calculation and render and has stricter requirements for hardware.

4.5.2 VR Live Broadcast Network Solution

Live broadcast by unicast has few fundamental difference with on-demand, and therefore see the above description for relevant information. For live broadcast of FOV VR videos by multicast, each program contains video stream of several views with each view corresponding to a multicast address. Client’s choosing of different views for display is to choose different multicast addresses.

Similar to on-demand, server needs to maintain views files and the corresponding angle relations of each program for live broadcast. The difference is the sites accesses is multicast address for live broadcast by multicast. When on-demand starts, client will request to server for program information to acquire list of multicast address and mutual angle relations.

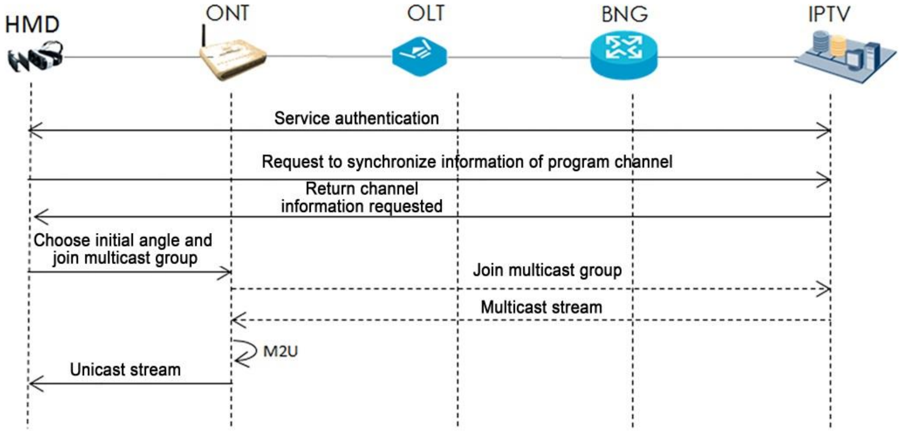
Table 4-4 Information on views of FOV VR live broadcast

Program	Multicast Address	Angle	Whether Initial View
Program a	235.254.196.1	0	**Yes**
	235.254.196.2	+30	
	235.254.196.3	+60	
	……	……	
	235.254.196.12	+330	
Program b	235.254.197.1	0	**Yes**
	235.254.197.2	+30	
	235.254.197.3	+60	
	……	……	
	235.254.197.12	+330	

The following figure shows the process of VR live broadcast initially joins multicast where ONT serves as IGMP Proxy. As a matter of fact, BNG and OLT can also serve as IGMP Proxy. Since the VR video stream is relatively large reaching Gbit/s level, the position of IGMP Proxy is suggested to be as close as possible to users to reduce broadband pressure of network to the minimum level.

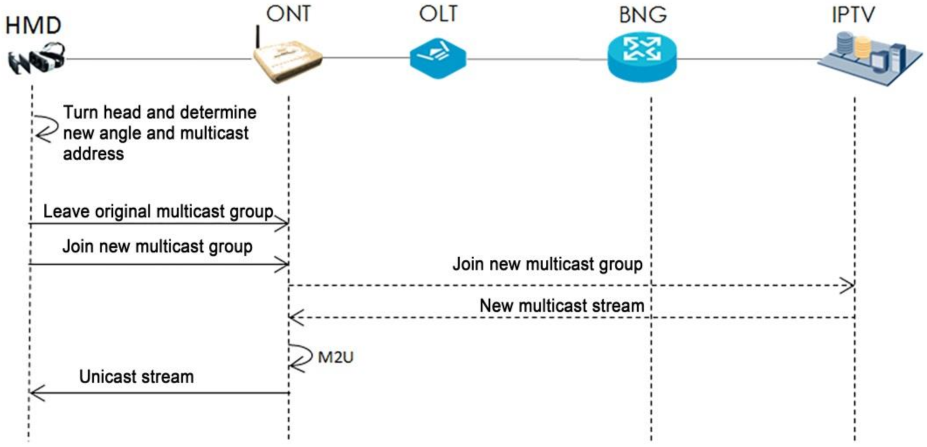
When client initially request for program information, it acquires the list of multicast address and angle information of the program and the initial view of multicast address suggested by sever. When choosing the initial content to display, client can use the multicast address of initial view given by server or choose on its own. Although the initial multicast address may be determined by client, once the initial view is chose, client shall use current view as reference and establish corresponding relations between local views and multicast addresses according to that of downloaded views and addresses from server so that when views changes due to head turning, client can switches to corresponding multicast.

Figure 4-7 VR live broadcast joins multicast



The process of client joining multicast after acquiring multicast address remains unchanged. If head does not turn or the turning angle does not exceed certain threshold during the process of display, the display of the video stream will continue with certain contents displayed with lower definition. When the turning angle of head exceeds threshold, since the view is diverged from the high definition area of former FOV, the definition of displayed content will reduce and therefore it is required to switch video stream to the multicast of new view.

Figure 4-8 Switch of angle of FOV VR live broadcast



In addition, when viewing traditional videos by IPTV, since users are “off” the scene, they have the preparation for picture switch and has little requirements for switching time of multicast. However, due to such unique features of VR as immersion and on the scene, the requirements for view switching time when turning head is very strict. Otherwise it is easy for users to sense the incoherence of pictures and feel uncomfortable due to the decreased resolution. If the 2s switching time of former IPTV channel is acceptable, the current switching time of VR displaying may be only one-tenth of the original time. Since multicast data is transmitted by constant bit rate, to accelerate switching speed, client may use unicast to download required data of new view to local at higher rate and display to achieve the best effect of video switching without been sensed by user when he/she turns head. However, one more process of frame stitching from unicast data to multicast data is added in such case.

4.6 Network with Customized Services

4.6.1 Current Network Status and Plights of Operators

There is an increasing amount of internet applications and new applications keep coming forth. Different applications require different network capacity to go with such application due to their features. Among these applications, video applications take a share of internet traffic for more than 70% and such share will increase in the future. With the development of videos, from 360p, 720p, 1080p to today’s 4K or even 8K, the requirement of visual experience become increasingly demanding to the sensibility limit of human eyes. However, the pursuit of sensory experience is endless. The emergence of virtual reality changes passive viewing to active participation when viewing videos. Compared with traditional “visual& auditory sense”, virtual reality combines such multidimensional information as visual, auditory and tactile feedback, and enhance from 2D image/video to 3D holographic image/video to provide multidimensional interactive experience from direct visual and auditory experience.

The initial goal of current IP-based internet design is to provide best data transfer. However, with the increasing amount of applications and various requirements for data transmission, the current QoS mechanism provides no difference transmission for the majority of data among the flood of internet other than the service with certain quality assurance for a few specific applications. As a service taking up an dominate share of internet traffic videos have relatively high requirements for bandwidth and latency. However the majority of videos traffic has no difference with regular internet traffic in practices. In such case, the transmission quality of videos are tempted to be affected resulting to poor experiences. In the process, operators are merely the channel for date transmission. Under the trend of increased speed and reduced cost of internet, on one hand, the price of broadband reduces continuously; on the other hand, network infrastructure investment by operators shall continue to increases. However the effect is not obvious, and people are still no satisfied. Over time, operators may lack sufficient power to continue expansion of network infrastructure. So far, the simple traffic-based charging mode will not increase income of operators and the unified charging rate by traffic for all OTT services will restrict internet innovation and therefore affects the flourish development of the entire internet. According to investigation, 85% of users are willing to pay at most 25% premium for better service experience. In other words, if operators can provide certain users with services of better quality, such users are willing to pay the cost of such high-quality services. For example, Netflix agreed to pay Comcast, an American operator to ensure that their users will have faster net speed in 2014. Therefore, operators shall turn the valuable network resources into a truly operational product rather than using as a signal transmission channel with no difference.

4.6.2 Service-customizable Network

Virtual reality has pushed people's experience to a new height thanks to its sense of immersion. At the same time, however, this sense of immersion brings a picture equal to a 1050-inch screen being watched from 3 meters away, which has advanced higher requirements for the resolution, refresh rate and latency of videos. Under virtual reality, to reach a maximum resolution of human eyes, the bandwidth required for VR panoramic video display will reach 2.35G (FOV) or even 5G (full perspective). End-to-end quality of such elephant flow with extremely high requirements cannot be ensured once there is a blockage in any node if the videos are only displayed with the contents provided by video content providers and via the pipes provided by operators and transmitted through normal data. There are two ideas to solve this problem: The first one is the network ensures the service as needed by perceiving the service features and transmission demands it bears; the second one is the network does not perceive the service, but the party where the service belongs or the user of the service applies resources that meet the transmission demands from the network. Because of the diversity and mutability of services, perceiving all service characteristics will result in complication of network and strong coupling with the service, adverse to the development of both service and network. The second idea is more appropriate, in which the pipe provider, namely the operator provides different services that are customizable as needed. This kind of services meet the requirement of on demand, being dynamic, being opening and end to end.

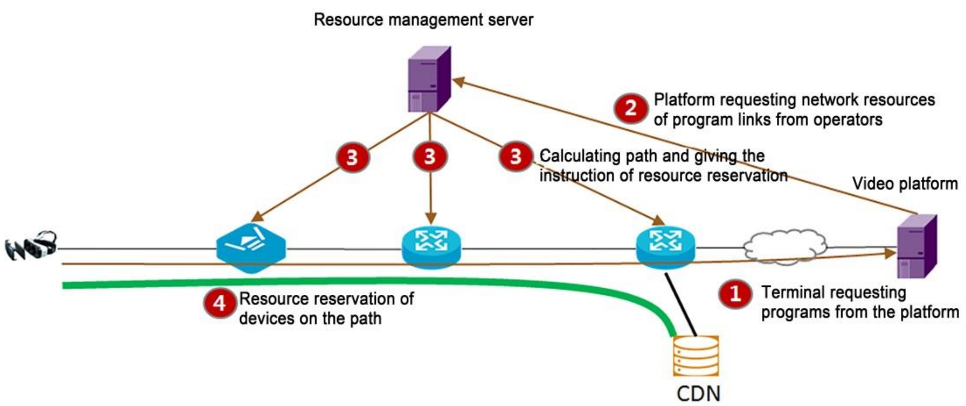
On demand: This includes two aspects. The first is the network requirements of the specific service, such as bandwidth, latency, packet loss rate and jitter. The network is able to identify the traffic/service object based on the source address, target address, quintuple or other service characteristics, meeting the traffic requirements of the object; the second is in terms of the network itself. With the increase of services, network resources should be capable of dynamic expansion according to service requirements. Otherwise, when service decreases, network resources should be reduced.

Dynamic: Because network requirements are different for videos and other services requiring quality assurance and there is uncertainty in times of occurrence and duration, QoS of network devices is not statically configured and reserved in advance but is computed end to end in seconds when the dialogue occurs, with resources allocated and scheduled at every node along the path. Resources stop releasing when the service terminates and are provided for other services to improve utilization.

Open: The operator provides amicable, specific and complete interfaces and customization of network quality for video services as required by OTT, including service application, adjustment, release, billing, account checking and settlement.

End-to-end: Service data flow serially passes every node on the internet. Blockage in any node on the path will influence the whole service. Therefore, to assure service quality, uniform management and calculation needs to be carried out end to end to ensure that each node provides sufficient quality assurance for the service. This requires a centralized management unit to obtain in real time the status (bandwidth, latency and jitter) of each device on the network bearing the service. When service request occurs, the unit can calculate a reasonable path according to the status of each device and give the instruction of resource reservation to each device on the path, setting the whole path ready.

Figure 4-9 Customization-based network service



Note that flow-based allocation of network resources is more appropriate for services that will constantly use network resources once occurring, such as video resources. In addition, the video source must ensure that the code rate it sends is within the scope of applied network resources, or otherwise the ultimate experience cannot be guaranteed.

Currently, there has already been some technical studies concerning customization of network resources based on services, such as vNet and DetNet. Here are some brief introduction to these two technologies.

1. **vNet (virtual network)**

   The principle of vNet is dynamically creating private logic networks according to the different characteristics of services. It mainly includes the following aspects:

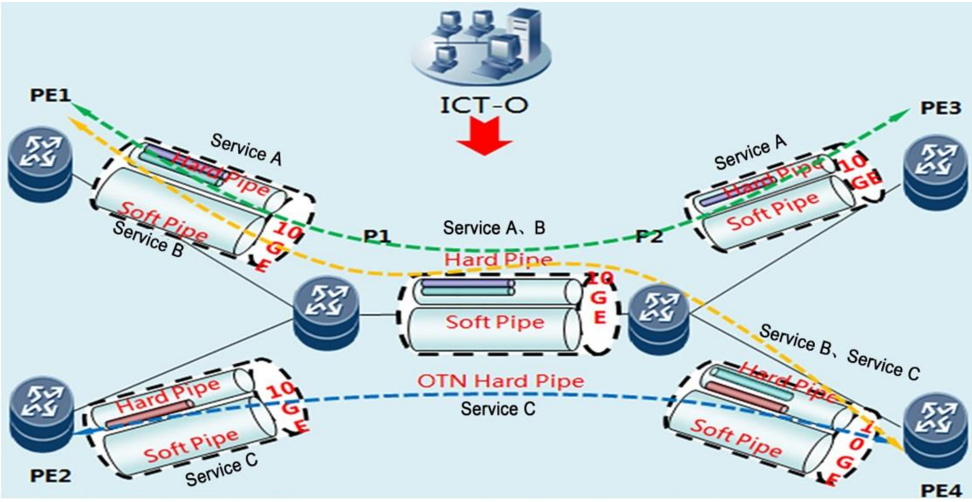
   (1). Service collaboration: service orientation, capability opening and slicing collaboration

   (2). Slicing management: life period management of network slicing

   (3). Resource management: virtual resource management and unified service distribution and maintenance based on Information And Communication Technology (ICT)

   (4). Network management: resource reservation and user plane isolation based on demands of virtual resources

Figure 4-10 vNet private logic network



2. **DetNet (determined network)**

   DetNet aims at providing service stream specific services which are based on layer 2 or layer 3 and feature limited delay, low packet loss rate, low jitter and high reliability and transferring the private networks of different areas into packet switched network. At the same time, DetNet integrates the TSN (Time Sensitive Network) based on operation of layer 2 and expands the common architecture for layer 2 and layer 3. The main technologies include:

   (1). Resource reservation: The application initiates the request as needed to reserve resources at every node passed by the flow, such as bandwidth and buffer. In this way, failure in property assurance caused by blockage will be eliminated. At the same, time-based resource sharing is needed among DetNet flows with high requirements of synchronism.

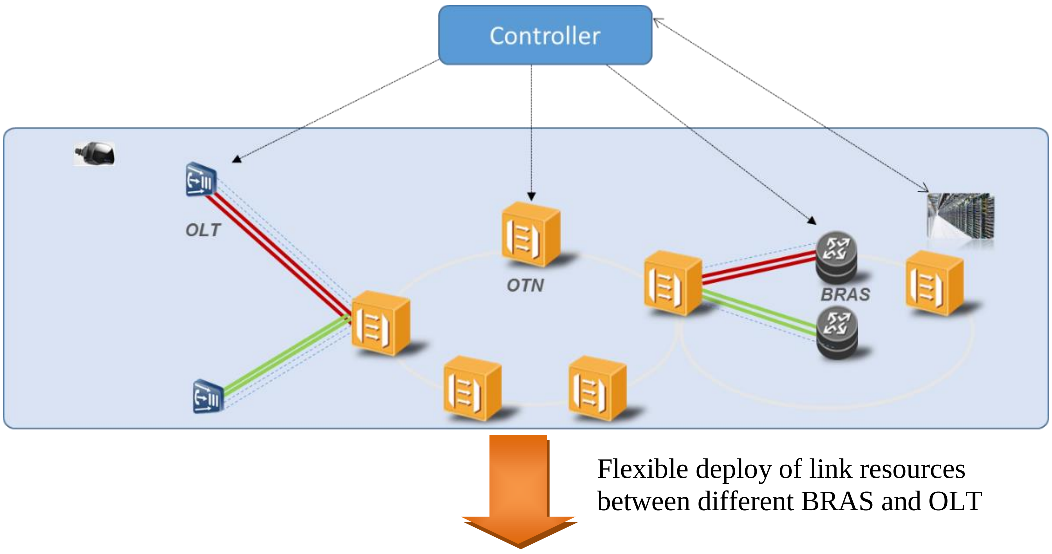
(2). Specifying path: The path of DetNet flow will not change easily even in face of emergencies of network. The middle transmission nodes requires no additional treatment, which reduces transmission time. The path employs MPLS TE LSP.

(3). Packet loss mechanism: In case of device failures on the path of packet transmission, copy the packets and transmit via multiple paths. Sort the packets at the receiving end and eliminate repeated packets.

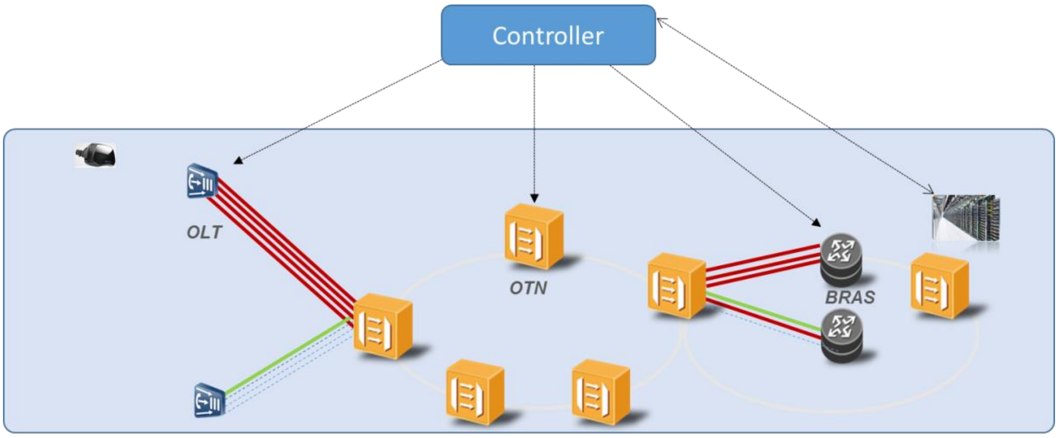
4.7 Network Dynamic Self-adaption – NDSA

Bearer network, to adapt to the great tide of VR service, must consider a new plan to meet service requirements and reduce CAPEX and OPEX of network. A basic idea is temporarily closing or releasing resources to other services when in low ebb, including links, wavelength division and forwarders: Closing resources to reduce power consumption of devices and further to reduce OPEX; or releasing resources to other services to reduce the total CAPEX.

The distinct difference between dynamics of this tide and that of the previous Internet or common HD videos is that, gap between the peak and valley is larger and the peak exits longer. If there is a 10-time gap between peak and valley of previous video services, there is possibly 100-time gap between peak and valley of VR service. Likewise, dynamic application and release of the network are likely to bring the operator's network greater value and the operator more cost savings.



Flexible deploy of link resources between different BRAS and OLT

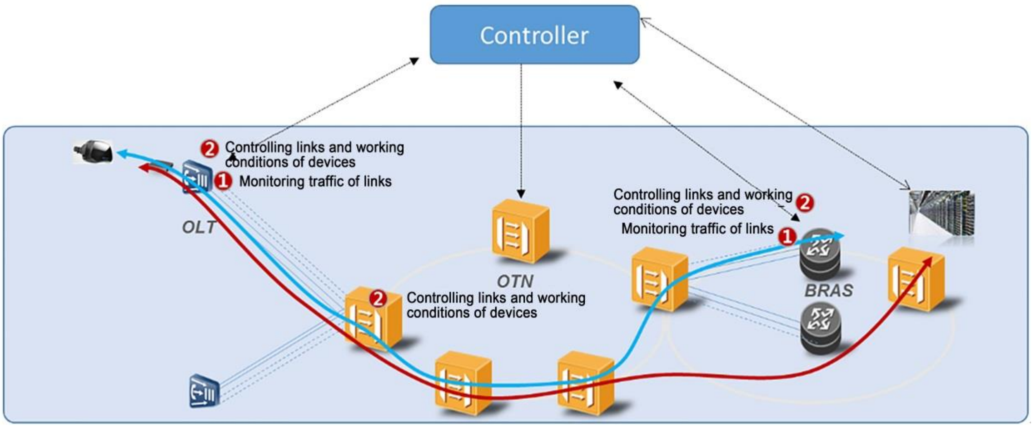


The following introduces three plans according to differences among methods and timings of resource dynamic application:

4.7.1 Dynamic Adjustment Based on Threshold of Link Utilization

The controller constantly acquires link utilization of devices and begins to notify relevant devices about the reserved resources to start pooling, including transmission resources, ports of forwarders and forwarding boards when detecting that some vital link resource or device resource reaches the set threshold. For NFV devices, more VMs need to be applied to join the forwarders to better cope with the upcoming peak flow and provide the best VR service experience. Likewise, the controller may temporarily close some redundant links, ports and boards and release some VMs to decrease power consumption, or release the useless resources to other services when detecting that the link flow is below the threshold for a continuous period (such as 10 min). In other words, the whole process is dynamic and on demand.

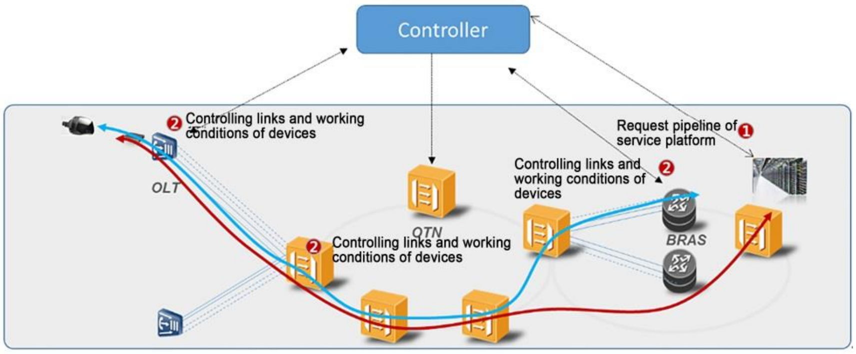
Figure 4-11 Dynamic adjustment based on threshold of link utilization



4.7.2 On-demand Service Application

When a VR user clicks a VR video, the platform providing the VR service will first acquire the information. The service platform then needs to send a request to the controller of the network and also information of the user. The controller will, based on the user's information (generally the addresses of his/her source IP and the service platform and code flow of the service), restores the path end to end, examines link resources on the whole forwarding path and utilization of forwarding resources and adds the upcoming VR service. If the service and bandwidth requirements of the user are well satisfied, corresponding links and forwarding resources needs to be applied in advance.

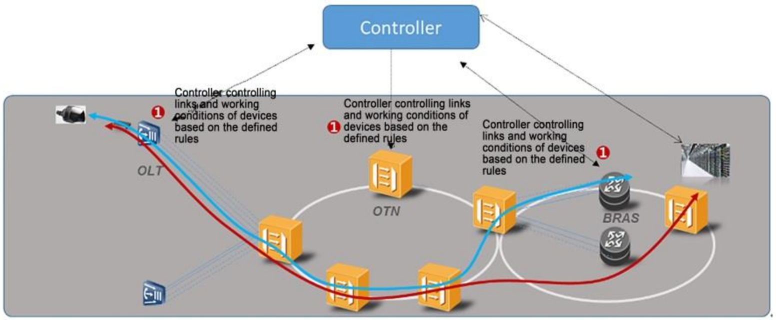
Figure 4-12 Service on-demand application



4.7.3 Time-based Dynamic Service Adjustment

We know that every service has its tidal pattern. For example, 20 ~ 22 o'clock is the peak hours of every day. For another example, rules can be found in the possibility of a current hot issue like a sports game. Likewise, similar rules can be found in the VR service and resources and application can be prepared in advance to ensure experience of the VR service.

Figure 4-13 Dynamic adjustment based on rule of time



Of course, the above plans are not mutually exclusive in specific implementation and application and can be deployed together with each other as needed. On basis of ensuring the service experience to the maximum, reduce network OPEX as far as possible and avoid over complexity of plan deploy. In addition, ensure that there is not too much cost incurred from the deploy and obtain a good balance between cost and complexity.

4.8 OTT VR Live Broadcast Solution

From 4K video services, the demand of network bandwidth has constantly increased as a result of code flow increase of services. It seems that the traditional solution of using TCP unicast for live broadcast, especially live broadcast of hot issues is hard to continue. One single current is for one user, and the total network bandwidth required in the whole live service is: number of users * bandwidths of a single current. When there is a hot issue, the number of online users constantly breaks the record, plus the greater bandwidth consumption of the single current of 4K videos, OTT live broadcast is in face of great challenge. For OTT content providers, more online users means more earnings (including user payments and advertisements). Although videos with greater code current like 4K videos may attract more users fond of high-quality videos, they will increase the cost for OTT at the same time. To ensure the user experience at peak hours, OTT content providers must buy network output bandwidth from operators as per the maximum number of online users when there is a hot issue, and even the phenomenon that traffic is increased but contribution to operators' revenue is not increased may occur. Of course, to increase the maximum capacity of online users, some OTT providers decreases code stream of the service when a hot issue is coming. However, this method is a double-edged sword. Decrease of code stream, while lowering the requirements for network, also reduces user experience, which is adverse to the long-term user development.

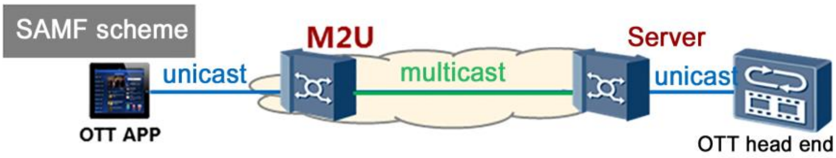
VR videos is more remarkable regarding this issue, compared with 4K HD videos. If OTT content providers can barely afford the cost pressure caused by HD videos, the super elephant flow caused by live broadcast of VR videos and the bandwidth consumption caused by the increase of online users are astonishing. If the bandwidth consumption for a user watching VR videos is 1 Gbit/s, the consumption is 100 T for 0.1 million users and 10000 T for 10 million users. Almost no OTT providers can bear such huge bandwidth, and even operators' networks cannot meet the demand in a short time. To solve the problems mentioned above, possible resolutions need to be considered.

4.8.1 Overview of SAMF Scheme

SAMF (Shared Adaptive Multicast Framework) scheme follows a simple principle. To reduce bandwidth occupation, live service traffic over unicast is converted into traffic over multicast and transmitted over the internet. From the aspect of multicast, bandwidth occupation is only related to the number of multicast channels but nothing to do with that of the online users. At the same time, as hot events are relatively centralized, the number of live channels is however not large. For the content provider of OTT VR live broadcast, CDN outlet bandwidth is no longer a bottleneck (one traffic for one channel, unrelated to the number of online users). It can greatly reduce the requirements of bandwidth at the network outlet to make the network bandwidth one hundred-thousandth or even of one ten-millionth of the original bandwidth and guarantee the televiewing experiences of users. Of course, as this method involves the change of business models, it needs the network provider to further open the network capacity, which can only be completed in operation with the OTT manufacturer.

To deploy this scheme from end to end, the operator shall first open its network to supply the OTT VR manufacturer with the capacity to apply network multicast (this capacity can only be applied in the self-supporting IPTV service of the operator at present).

Figure 4-14 Principle of SAMF scheme



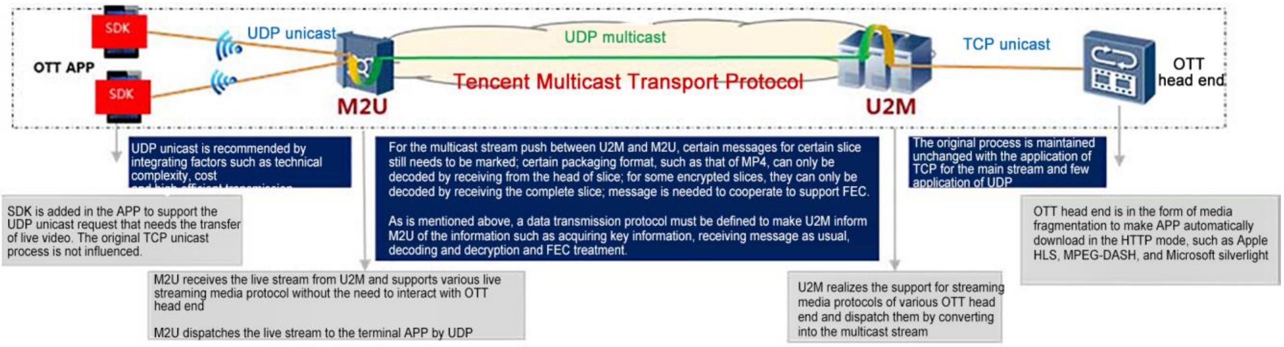
From the above figure, we can see that there is a big change in the whole traffic model from originally transmitting several traffics for several users to only transmitting one traffic (for one live channel). It greatly reduces the consumption of network bandwidth and implies the reduction of use-cost of network.

Unlike common TV large screens, this TV large screen's position is relatively fixed and its wired connection is able to meet the requirements. For VR glasses, wireless connection is a mode that is easier for the final user to accept. Conversion of multicast stream into unicast stream is needed because multicast is not friendly supported by Wi-Fi connection between the terminal and household gateway. (notes: A new Wi-Fi protocol may be more friendly for the support of multicast).

4.8.2 Workflow of SAMF Scheme

With SAMF, the traditional E2E unicast is converted into a three-stage architecture of unicast-multicast-unicast.

Figure 4-15 Process flow of SAMF Scheme



Stage I: between U2M equipment and OTT head end (unicast). U2M realizes the support for the streaming-media protocol at various OTT head ends and converts it into the dispatch of multicast stream. TCP unicast is applied between U2M and OTT head end. Firstly, the unicast is used for reducing the difficulty of butt joint of U2M equipment and OTT head end, and then the TCP/HTTP is for maintaining the format of media stream of original head end system.

Stage II: between U2M equipment and M2U equipment (multicast). The standard multicast forwarding is used between U2M equipment and M2U equipment. The benefits of the application of multicast is the core of this scheme, namely greatly reducing the consumption of network bandwidth and supporting the streaming-media protocol of various live broadcasting without the need of interaction of M2U and OTT head end.

Stage III: between M2U equipment and terminal APP (unicast). As is mentioned above, it is a good choice to dispatch M2U to the terminal APP in a unicast way because Wi-Fi cannot support the multicast flow very well. The specific application of UDP or TCP in the transport protocol depends on the comparison of the advantages and disadvantages of the two modes.

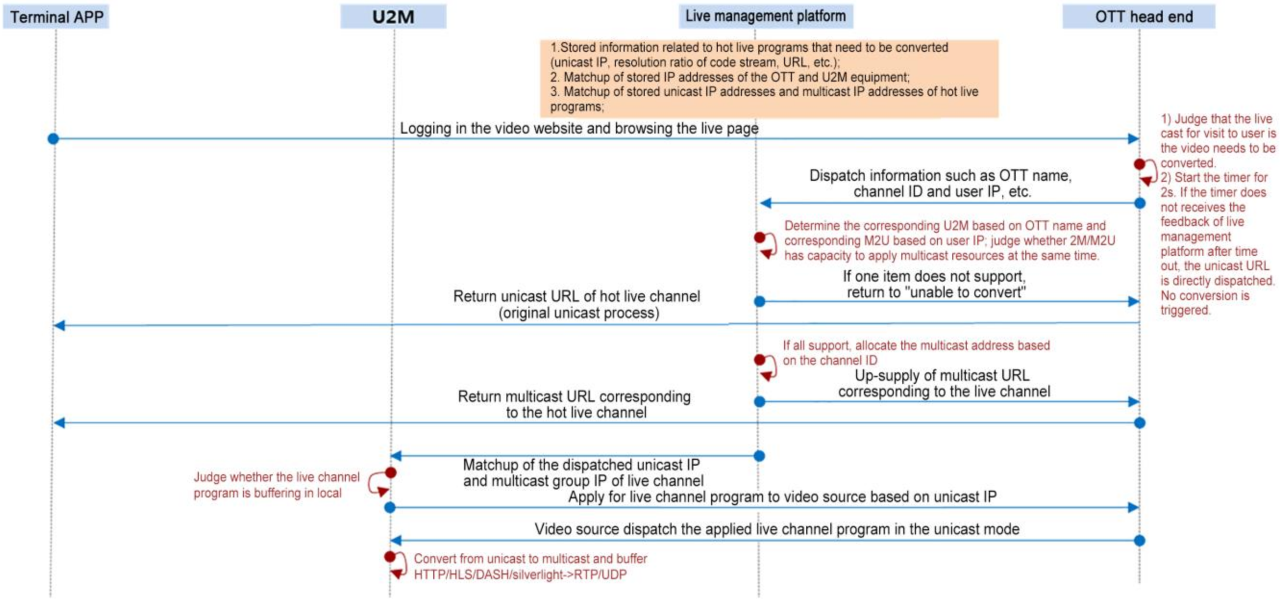
Table 4-5 Transport mode between M2U and terminal APP

Scheme comparison	Number of supported channels	Extension of streaming-media protocol	Restructuring difficulty	Advantage	Disadvantage
UDP mode	Bufferless and under the restriction of uplink bandwidth. 10Gbit/s uplink bandwidth can support 500 channels.	General. The UDP mode is relatively less, represented by GUIC.	Small. Our company provides the packaged UDP communication SDK and OTT APP only needs to call the standard SDK interface.	1. Supporting encrypted video streaming; 2. Small bandwidth requirements; 3. Strong performance and low latency of forwarding; 4. Forwarding with boards and low cost.	1. APP must be adapted and modified; 2. The packet loss and retransmission mechanism needs to be deployed.
TCP/HTTP Mode	Based on the storage limits of about 30, it is also under the bottleneck restriction of downlink bandwidth. Generally M2U for the most popular channel (<=10);	Convenient. All the mainstream live streaming-media protocols in the industry are based on HTTP at present.	Extremely small. Following the mainstream HLS downloading method now basically without any modification.	1. No modification of APP; 2. Supporting packet loss and retransmission. 3. No index files and fragmentation of M2U are generated and decryption of video streaming is not needed.	1. Restricted M2U performance and relatively low specification; 2. Complicated technology and excessive cost.

For the analysis of the above table, the application of UDP mode is the preferential choice of this scheme because it can effectively reduce the requirements on the network equipment to make the M2U equipment not understand and participate in the concrete service and involve in the treatment beyond the transport layer and reduce the CPU and storage requirements of the forwarding equipment. The disadvantage is that OTT APP needs to be adapted in accordance with the new protocol (SDK).

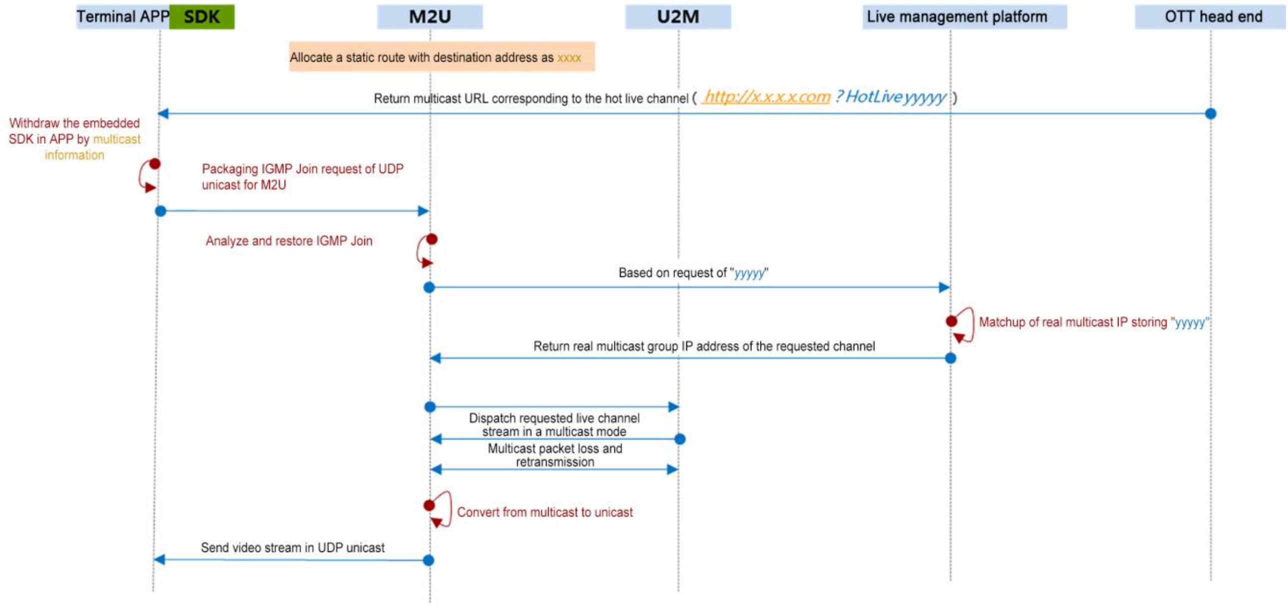
At the same time, as at least one complete sliced file and index file is needed during the broadcasting of APP (the following takes the most frequently used HLS of the OTT video), the first message received through UDP mode may not be the first message of one sliced file. At this moment, the condition that immediate playing is not available will happen and the next message of sliced file needs to be waited. This problem prolongs the time length from program selection by a user to start of playing (under the worst conditions, the additional time for waiting one time slice (TS) is 10s by default). Of course, the scheme can consider to transmit the first time slice by applying the original unicast process and switch to the multicast mode after waiting for the multicast mode completely to receive one time slice. By this way, the waiting time of user can be greatly reduced (Note: A complete scheme will involve in the butt joint of BSS/OSS system, so this white paper only discusses the working process of service).

Figure 4-16 Treatment process of the U2M services



Treatment process of the U2M services: first support the terminal APP of new SDK to browse the program (acquiring the program is the original process). After the user selects the confirmed live channels, OTT head end will interact with the live platform according to the selected live channel (the process of on-demand unicast is not changed): send the information such as OTT name, channel ID and user's IP to the live platform. The live platform feeds back the multicast URL of the OTT head end and sends the unicast IP address and multicast group address of program source to U2M. U2M interacts with the OTT head end with unicast IP to acquire and store the video content the stored contents are sent in a multicast way.

Figure 4-17 Treatment process of the M2U services



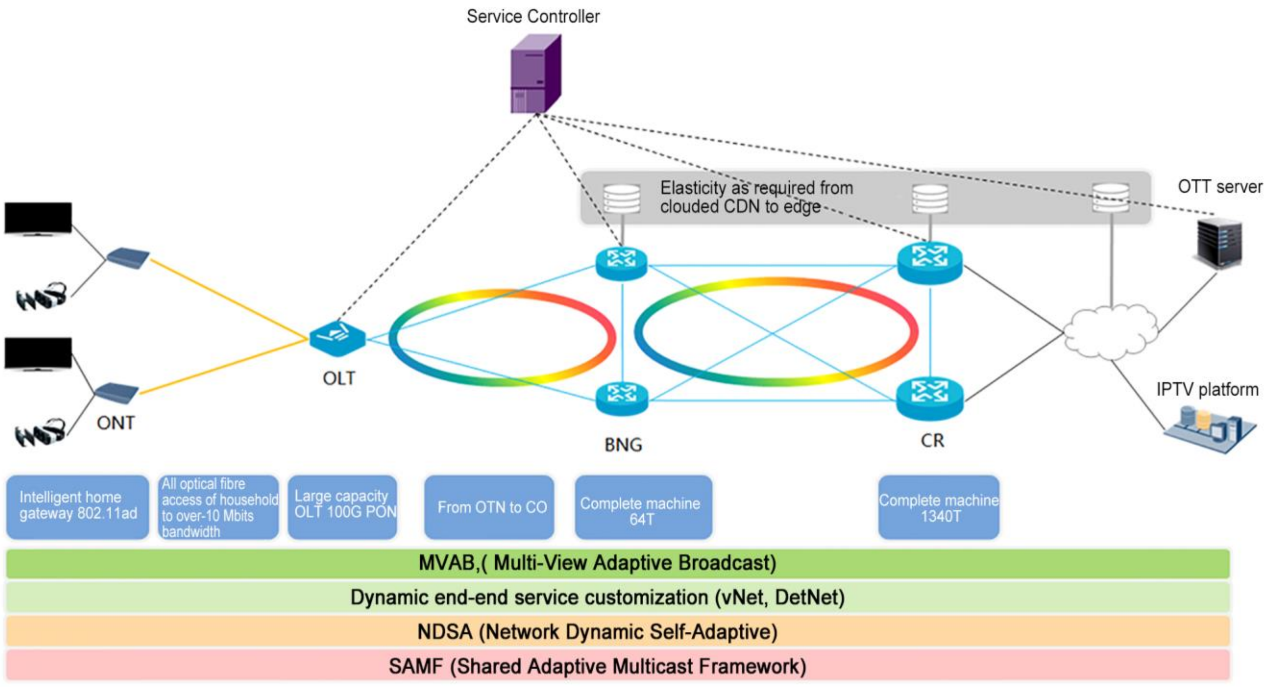
Treatment process of the M2U services: The terminal APP takes the multicast URL returned from OTT head end as the SDK function for parameter calling. The SDK function will analyze the inner multicast information and sends it to the M2U equipment. According to the received information, the M2U equipment request the multicast address to the live platform and joins in the corresponding multicast group to achieve the video content in multicast mode. The M2U equipment will send the acquired video content to the terminal APP in a unicast mode.

4.9 Effect Analysis of Network Architecture

As is mentioned above, the development of VR services has a significant effect on the network, leading to the corresponding changes in the network architecture. These changes are mainly embodied in the following aspects:

1. Ultra high bandwidth;
2. Dynamic adaptive network capacity;
3. Further flexible and opening network capacity (such as multicast);
4. Coordinating with services to make network no longer a black box to the services.

Figure 4-18 VR Oriented Bearer Network



The development of VR also follows the rule that the things develop in an evolutionary way. This rule applies from the entry-level experience stage to advanced experience stage, and then to extreme experience stage, from penetration and concurrency rate of users to the resolution, refresh rate, color depth of VR video and with the improving of encoding and decoding technology. Likewise, the effect of VR on the network architecture also smoothly develops step by step. With the changing of the requirement of this process on the bearer network, the network will gradually evolve too. From the above described features, the differences in the effect and importance of VR in different stages can be gradually reflected in the network architecture.

5 Development of VR Application Services and Expectation of Future Network

As a brand new platform, in addition to the entertainment direction such as panoramic videos and games, VR is destined to develop more applications facing the industries, such as shopping, house-inspecting, travel and education. VR will bring about more innovation of the business model as the same with the “position sponsoring” from Pokemon Go. There is a great market space for these industries, which is destined to provide new opportunities to practitioners. Therefore, VR services are one of the strategic peaks facing the future. All the far-sighted practitioners in these industries realize this opportunity and take VR as an important business strategy.

Looking back from the future, VR is destined to gradually tend to the integration, MR along with AR (augment reality). AR and MR will bring about new requirements and challenges to the network. In addition to the introduction of more information amount, another change in AR and MR is that more calculation is completed on the cloud. One of the conveniences for this change is that the network is required to provide lower latency and higher bandwidth, especially the uplink bandwidth; and the other convenience is that the network is required to be more coordinated and integrated with the cloud to become the “intelligent pipe” for storage, calculation and information transmission as required. More requirements are proposed for large bandwidth and network efficiency, especially the energy efficiency ratio. It is especially important to continuously reduce the transmission cost per bit.

In a word, a network with large bandwidth, low latency and high energy efficiency needs to be provided for VR and new requirements for the on-demand capacity of the network is proposed at the same time. The progress of basic communication technology and rapid development of SDN/NFV architecture lay a solid foundation for the bearer network facing the VR experiences.

The user’s pursuit to experiences will never be an end. This is fully manifested in the development of ordinary video from SD to 4k. We believe that VR will soon enter the “2K” or even “4K” stage. In a sense, VR products in the current stage fast cultivate the habits of users and popularize the recognition of consumers on VR services, resulting in certain positive effect on the development of industries. However, we shall also realize that real VR can do better. Therefore a standardized evaluation system based on experiences is needed urgently in the VR industry. This system is not only conducive to promoting the growth of VR products but also leading the industry to rapidly making up the weak points of various links such as content, auxiliary products, and bearer network, and increasing the popularization of VR application. Huawei is conducting the relevant research and expecting to jointly promote the maturity of standards and development of industry with partners in the industrial chain and make a contribution to the establishment of a better fully-connected world.