L. Han, Ed.
Huawei Technologies
S. Appleby
BT
K. Smith
Vodafone
March 12, 2017

Problem Statement: Transport Support for Augmented and Virtual Reality Applications
draft-han-iccrg-arvr-transport-problem-00

Abstract

As emerging technology, Augmented Reality (AR) and Virtual Reality (VR) bring up a lot of challenges to technologies such as information display, image processing, fast computing and networking. This document will analyze the requirements of AR and VR to networking, especially to transport protocol.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction 2
  1.1. Scope 3
2. Terminology 4
  2.1. Definitions 4
3. Problem Statement 6
4. IANA Considerations 9
5. Security Considerations 10
6. Acknowledgements 10
7. References 10
  7.1. Normative References 10
  7.2. Informative References 10
Appendix A. Key Factors for Network-Based AR/VR 12
  A.1. Latency Requirements 12
    A.1.1. Motion to Photon (MTP) Latency 12
    A.1.2. Latency Budget 13
  A.2. Throughput Requirements 15
    A.2.1. Average Throughput 15
    A.2.2. Peak Throughput 19
Authors' Addresses 20

1. Introduction

Virtual Reality (VR) and Augmented Reality (AR) technologies have enormous potential in many different fields, such as entertainment, remote diagnosis, or remote maintenance. AR and VR applications aim to cause users to perceive that they are physically present in a non-physical or partly non-physical world. However, slightly unrealistic artefacts not only distract from the sense of immersion, but they can also cause 'VR sickness' [VR-Sickness] by confusing the brain whenever information about the virtual environment is good enough to be believable but not wholly consistent.

This document is based on the assumption and prediction that the current localized AR/VR will inevitably evolve to cloud based AR/VR. Since cloud processing and state will be able to supplement local AR/VR devices, helping to reduce their size and power consumption, and to provide much more content resource and flexibility to the AR/VR applications.

Sufficient realism requires both very low latency and a very high information rate. In addition the information rate varies significantly and can include large bursts. This problem statement aims to quantify these requirements, which are largely driven by the video component of the transmission. The ambition is to improve Internet technology so that AR/VR applications can create the impression of remote presence over longer distances.

The goal is for the Internet to be able to routinely satisfy these demanding requirements in 5-10 years. Then it will become feasible to launch many new applications, using AR/VR technology in various arrangements as a new platform over the Internet. A 5-10-year horizon is considered appropriate, given it can take 1-2 years to socialize a grand challenge in the IRTF/IETF then 2-3 years for standards documents to be drafted and taken through the RFC process. The technology itself will also take a few years to develop and deploy. That is likely to run partly in parallel to standardization, so the IETF will need to be ready to intervene wherever interoperability is necessary.

1.1. Scope

This document is aimed at the transport area research community. However, initially, advances at other layers are likely to make the greatest inroads into the problem, for example:

o Network architecture: the physical distance between the content cloud of AR/VR and users are short enough to limit the latency caused by the propagation delay in physical media

o Motion sensors: reduction in latency for range of interest (RoI) detection

o Sending app: better targeted degradation of quality below the threshold of human perception, e.g. outside the range of interest

o Sending app: better coding and compression algorithms

o Access network: multiplexing bursts further down the layers and therefore between more users, e.g. traffic-dependent scheduling between layer-2 flows not layer-3 flows

o Core network: The capacity of the core network is sufficient to support transport of AR/VR traffic cross different service providers.

o Receiving app: better decoding and prediction algorithms

o Head mounted displays (HMDs): reducing display latency

The initial aim is to state the problem in terms of raw information rates and delays. This initial draft can then form the basis of discussions with experts in other fields, to quantify how much of the problem they are likely to be able to remove. Then subsequent drafts can better quantify the size of the remaining transport problem.

This document focuses on unicast-based AR/VR, which covers a wide range of applications, such as VR gaming, shopping, surgery, etc. Broadcast/multicast-based AR/VR is outside the scope of this document. It is likely to need more supporting technology such as multicast, caching and edge computing. Broadcast/multicast-based AR/VR is for live or multi-user events, such as sports broadcasts or online education. The idea is to use panoramic streaming technologies such that users can dynamically select different view points and angles to become immersed in different real time video streams.

Our intention is not to promote enhancement of the Internet specially for AR/VR applications. Rather AR/VR is selected as a concrete example that encompasses a fairly wide set of applications. It is expected that an Internet that can support AR/VR will be able to support other applications requiring both high throughput and low latency, such as interactive video. It should be able to support applications with more demanding latency requirements, but perhaps only over shorter distances. For instance, low latency is needed for vehicle to everything (V2X) communication, for example between vehicles on roads, or between vehicles and remote cloud computing. Tactile communication has very demanding latency needs, perhaps as low as 1 ms.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2.1. Definitions

E2E
: End-to-end

HMD
: Head-Mounted Display or Device

AR
: Augmented Reality (AR) is a live direct or indirect view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as

sound, video, graphics or GPS data. It is related to a more general concept called mediated reality, in which a view of reality is modified (possibly even diminished rather than augmented) by a computer

VR
: Virtual Reality (VR) is a computer technology that uses software-generated realistic images, sounds and other sensations to replicate a real environment or an imaginary setting, and simulates a user’s physical presence in this environment to enable the user to interact with this space

FOV
: Field of View is the extent of the world that is visible without eye movement, measured in degrees of visual angle in the vertical and horizontal planes

Panorama
: Panorama is any wide-angle view or representation of a physical space, whether in painting, drawing, photography, film, seismic images or a three-dimensional model

360 degree video
: 360-degree videos, also known as immersive videos or spherical videos, are video recordings where a view in every direction is recorded at the same time, shot using an omnidirectional camera or a collection of cameras. Most 360-degree video is monoscopic (2D), meaning that it is viewed as a one (360x180 equirectangular) image directed to both eyes. Stereoscopic video (3D) is viewed as two distinct (360x180 equirectangular) images directed individually to each eye. 360-degree videos are typically viewed via personal computers, mobile devices such as smartphones, or dedicated HMD

MTP and MTP Latency
: Motion-To-Photon. Motion-to-Photon latency is the time needed for a user movement to be fully reflected on a display screen [MTP-Latency].

Unmanaged
: For the purpose of this document, if an unmanaged Internet service supports AR/VR applications, it means that basic connectivity provides sufficient support without requiring the application or user to separately request any additional service, even as a once-off request.

3. Problem Statement

Network based AR/VR applications need both low latency and high throughput. We shall see that the ratio of peak to mean bit-rate makes it challenging to hit both targets. To satisfy extreme delay and throughput requirements as a niche service for a few special users would probably be possible but challenging. This document envisages an even more challenging scenario; to support AR/VR usage as a routine service for the mass-market in the future. This would either need the regular unmanaged Internet service to support both low latency and high throughput, or it would need managed Internet services to be so simple to activate that they would be universally accessible.

Each of the elements of the above requirements are expanded and quantified briefly below. The figures used are justified in depth in Appendix A.

MTP Latency: AR/VR developers generally agree that MTP latency becomes imperceptible below about 20 ms [Carmack13]. However, some research has concluded that MTP latency MUST be less than 17ms for sensitive users [MTP-Latency-NASA]. Experience has shown that standards bodies tend to set demanding quality levels, while motivated humans often happily adapt to lower quality although they struggle with more demanding tasks. Therefore, we MUST be clear that this 20 ms requirement is designed to enable immersive interaction for the same wide range of tasks that people are used to undertaking locally.

Latency Budget: If the only component of delay was the speed of light, 20 ms round trip would limit the physical distance between the communicating parties to 3,000 km in air or 2,000 km in glass. We cannot expand the physical scope of an AR/VR application beyond this speed-of-light limit. However, we can ensure that application processing and transport-related delays do not significantly reduce this limited scope. As a rule of thumb they should consume no more than 5-10% (1-2 ms) of this 20 ms budget, and preferably less. See Appendix A.1 for the derivation of these latency requirements.

	Entry-level	Advanced	Ultimate 2D	Ultimate 3D
Video Type	4K 2D	12K 2D	24K 2D	24K 3D
Mean bit rate	22 Mb/s	400 Mb/s	2.9 Gb/s	3.3 Gb/s
Peak bit rate	130 Mb/s	1.9 Gb/s	29 Gb/s	38 Gb/s
Burst time	33 ms	17 ms	8 ms	8 ms

Table 1: Raw information rate requirements for various levels of AR/VR (YUV 420, H.265)

Raw information rate: Table 1 shows the summary of mean and peak raw information rate for four types of H.265 video. Not only does the raw information rate rise to very demanding levels, even for 12K 'Advanced AR/VR'. But the ratio of peak to mean increases from about 6 for 'Entry-Level' AR/VR to nearly 12 for 'Ultimate 3-D' AR/VR. See Appendix A.2 for more details and derivation of these rate requirements.

Buffer constraint: It will be extremely inefficient (and therefore costly) to provide sufficient capacity for the bursts. If the latency constraint were not so tight, it would be more efficient to provide less capacity than the peak rate and buffer the bursts (in the network and/or the hosts). However even if capacity were only provided for 1/k of the peak bit rate, play-out would be delayed by (k-1) times the burst time. For instance, if a 1G b/s link were provided for 'Advanced' AR/VR, we can see that k = 1.9. Then play-out would be delayed by (1.9 - 1) * 17 ms = 15 ms. This would consume 75% of our 20 ms delay budget. Therefore, it seems that capacity sufficient for the peak rate will be needed, with no buffering. We then have to rely on application-layer innovation to reduce the peak bit rate.

Simultaneous bursts: One way to deal with such a high peak-to-mean ratio would be to multiplex multiple AR/VR sessions within the same capacity. This problem statement assumes that the bursts are not correlated at the application layer. Then the probability that most sessions burst simultaneously would become tiny. This would be useful for the high degree of statistical multiplexing in a core network, but it would be less useful in access networks, which is where the bottleneck usually is, and where the number of AR/VR sessions in the same bottleneck might often be close to 1. Of course, if the bursts are correlated between different users, there will be no multiplexing gain.

Problems with Unmanaged TCP Service: An unmanaged TCP solution would probably use some derivative of TCP congestion control [RFC5681] to adapt to the available capacity. The following problems with TCP congestion control would have to be solved:

Transmission loss and throughput: TCP algorithms collectively induce a low level of loss, and the lower the loss the faster they go. TCP throughput is used to measure such performance. No matter what TCP algorithm is used, the TCP throughput is always capped by some parameters, such as RTT, packet loss ration, etc. Importantly, the TCP throughput is always lower than the physical link capacity. So, for a single flow to attain the bit-rates shown in Table 1 requires a loss probability that is so low that it could be physically limited by the bit-error probability experienced over optical fiber links. The analysis [I-D.ietf-tcpm-cubic] has collected the data for different TCP throughput and corresponding packet loss ration.

Flow-rate equality:

Host-Controlled: TCP ensures rough equality between L4 flow rates as a simple way to ensure that no individual flow is starved when others are not [RFC5290]. Consider a scenario where one user has a dedicated 2 Gb/s access line, and they are running an AR/VR applications that needs a minimum of 400 Mb/s. If the AR/VR app used TCP, it would fail whenever the user (or their family) happened to start more than 4 other TCP long flows at once, i.e, FTP flows. This simple example shows that flow-rate equality will probably need to be relaxed to enable support for AR/VR as part of the regular unmanaged Internet service. Fortunately, when there is enough capacity for one flow to get 400 Mb/s, every flow does not have to get 400 Mb/s to ensure that no-one starves. This line of logic could allow flow-rate equality to be relaxed in transport protocols like TCP.

Network-Enforced: However, if parts of the network were enforcing flow rate equality, relaxing it would be much more difficult. For instance, deployment of the per-flow queuing scheduler in fq_CoDel [I-D.ietf-aqm-fq-codel] will introduce this problem.

Dynamics: The bursts shown in Table 1 would be problematic for TCP. It is hard for the throughput of one TCP flow to jump an order of magnitude for one or two round trips, and even harder for other TCP flows to yield over the same time-scale without considerable queuing delay and/or loss.

Problems with Unmanaged UDP Service: Using UDP as transport cannot solve the problems as faced by TCP. Fundamentally, IP network can only provide the best-effort service, no matter if the transport on top of IP is TCP or UDP. This is determined by the fact that most of network devices use different variations of "Fair Queuing" algorithm to queue IP flows without the awareness of TCP or UDP protocol. As long as a fair queuing algorithm is used, a UDP flow cannot obtain more bandwidth or shorter latency than others. But using UDP may reduce the burden of re-transmission of lost packet, if the lost packet is not so critical, like a non I-frame; or the lost packet has passed its life cycle. Depending on if it has its own congestion control, current UDP service has two types:

UDP with congestion control: QUIC is a typical UDP service with congestion control. The congestion control algorithm used in QUIC is similar to TCP CUBIC. This makes QUIC behave also similar to TCP CUBIC. There will be no fundamental difference compared with unmanaged TCP service in terms of fairness, convergence and bandwidth utilization, etc.

UDP without congestion control: If UDP is used as transport without extra congestion control, it will be weaker than with congestion control to support the AR/VR application with high throughput and short latency requirements.

Problems with Managed Service: As well as the common problems outlined above, such as simultaneous bursts, the management and policy aspects of managed QoS solution are problematic:

Complex provisioning: Currently QoS services are not straightforward to enable, which would make routine widespread support of AR/VR unlikely. It has proved particularly hard to standardize how managed QoS services are enabled across host-network and inter-domain interfaces.

Universality: For AR/VR support to become widespread and routine, control of QoS provision would need to comply with the relevant Net Neutrality [NET_Neutrality_ISOC] legislation appropriate to the jurisdictions covering each part of the network path.

4. IANA Considerations

There is no change with regards to IANA

5. Security Considerations

There is no security issue introduced by this document

6. Acknowledgements

Special thanks to Bob Briscoe, he has given a lot advice and comments during the period of study and writing of this draft, he also has done a lot revision for the final draft.

We would like to thank Kjetil Raaen for comments on early drafts of this work.

We also like to thank Huawei’s research team leaded by Lei Han, Feng Li and Yue Yin to provide the prospective analysis; also thank Guoping Li, Boyan Tu, Xuefei Tang and Tao Ma from Huawei for their involvement in the work discussion

Lastly, we want to thank Huawei’s Information LAB, the basic AR/VR data was from its research results

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

7.2. Informative References

[Carmack13] Carmack, J., "Latency Mitigation Strategies", February 2013, <https://www.twentymilliseconds.com/post/latency-mitigation-strategies/>.

[Chroma] Wikipedia, "Chroma subsampling", 2016, <https://en.wikipedia.org/wiki/Chroma_subsampling>.

[Fiber-Light-Speed] Kevin Miller, "Calculating Optical Fiber Latency", 2012, <http://www.m2optics.com/blog/bid/70587/Calculating-Optical-Fiber-Latency>.

[GOP] Wikipedia, "Group of pictures", 2016, <https://en.wikipedia.org/wiki/Group_of_pictures>.

[H264_Primer]
Adobe, "H.264 Primer", 2016, <http://wwwimages.adobe.com/content/dam/Adobe/en/devnet/video/articles/h264_primer/h264_primer.pdf>.

[I-D.ietf-aqm-fq-codel]
Hoeiland-Joergensen, T., McKenney, P., dave.taht@gmail.com, d., Gettys, J., and E. Dumazet, "The FlowQueue-CoDel Packet Scheduler and Active Queue Management Algorithm", draft-ietf-aqm-fq-codel-06 (work in progress), March 2016.

[I-D.ietf-tcpm-cubic]
Rhee, I., Xu, L., Ha, S., Zimmermann, A., Eggert, L., and R. Scheffenegger, "CUBIC for Fast Long-Distance Networks", draft-ietf-tcpm-cubic-04 (work in progress), February 2017.

[MTP-Latency]
Kostov, G., "Fostering Player Collaboration Within a Multimodal Co-Located Game", University of Applied Sciences Upper Austria, Masters Thesis , September 2015, <https://www.researchgate.net/publication/291516650_Fostering_Player_Collaboration_Within_a_Multimodal_Co-Located_Game>.

[MTP-Latency-NASA]
Bernard D. Adelstein, et al, NASA Ames Research Center, etc, "HEAD TRACKING LATENCY IN VIRTUAL ENVIRONMENTS: PSYCHOPHYSICS AND A MODEL", 2003, <https://humansystems.arc.nasa.gov/publications/Adelstein_2003_Head_Tracking_Latency.pdf>.

[NET_Neutrality_ISOC]
Internet Society, "Network Neutrality, An Internet Society Public Policy Briefing", 2015, <http://www.internetsociety.org/sites/default/files/ISOC-PolicyBrief-NetworkNeutrality-20151030-nb.pdf>.

[PSNR]   Wikipedia, "Peak signal-to-noise ratio", 2016, <https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio>.

[Raaen16] Raaen, K., "Response time in games : requirements and improvements", University of Oslo, PhD Thesis , February 2016, <http://home.ifi.uio.no/paalh/students/KjetilRaaen-phd.pdf>.

[RFC5290] Floyd, S. and M. Allman, "Comments on the Usefulness of Simple Best-Effort Traffic", RFC 5290, DOI 10.17487/RFC5290, July 2008, <http://www.rfc-editor.org/info/rfc5290>.

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", RFC 5681, DOI 10.17487/RFC5681, September 2009, <http://www.rfc-editor.org/info/rfc5681>.

[VR-Sickness] Wikipedia, "Virtual reality sickness", 2016, <https://en.wikipedia.org/wiki/Virtual_reality_sickness#cite_note-one-1>.

[YUV] Wikipedia, "YUV", 2016, <https://en.wikipedia.org/wiki/YUV>.

Appendix A. Key Factors for Network-Based AR/VR

A.1. Latency Requirements

A.1.1. Motion to Photon (MTP) Latency

Latency is the most important quality parameter of AR/VR applications. With streaming video, caching technology located closer to the user can reduce speed-of-light delays. In contrast with AR/VR user actions are interactive and rarely predictable. At any time a user can turn the HMD to any angle or take any other action in response to virtual reality events.

AR/VR developers generally agree that MTP latency becomes imperceptible below about 20 ms [Carmack13]. However, some research has concluded that MTP latency MUST be less than 17ms for sensitive users [MTP-Latency-NASA]. For a summary of numerous references concerning the limit of human perception of delay see the thesis of Raaen [Raaen16].

Latency greater than 20 ms not only degrades the visual experience, but also tends to result in Virtual Reality Sickness [VR-Sickness]. Also known as cybersickness, this can cause symptoms similar to motion sickness or simulator sickness, such as general discomfort, headache, nausea, vomiting, disorientation, etc.

Sensory conflict theory believes that sickness can occur when a user’s perception of self-motion is based on inconsistent sensory inputs between the visual system, vestibular (balance) system, and non-vestibular proprioceptors (muscle spindles), particularly when these inputs are at odds with the user’s expectations from prior experience. Sickness can be minimized by keeping MTP latency below the threshold where humans can detect the lag between visual input and self-motion.

The best localized AR/VR systems have significantly improved speed of sensor detection, display refresh, and GPU processing in their head-mounted displays (HMDs) to bring MTP latency below 20 ms for localized AR/VR. However, network-based AR/VR research has just started.

A.1.2. Latency Budget

Figure 1 illustrates the main components of E2E delay in network-based AR/VR.

```
+------+            +------+            +------+
|  T1  |----------->|  T4  |------------>|  T2  |
+------+            +------+            +------+
                                           |
                                           |
                                           |
+------+                                   |
|  T6  |                                   |
+------+                                   |
   ^                                       |
   |                                       |
   |                                       v
+------+            +------+            +------+
|  T5  |<-----------|  T4  |<------------|  T3  |
+------+            +------+            +------+

 T1:  Sensor detection and Action capture
 T2:  Computation for ROI (Range of Interest) processing, rendering
      and encoding
 T3:  GOP (group of pictures) framing and streaming
 T4:  Network transport
 T5:  Terminal decoding
 T6:  Screen refresh
```

Figure 1: The main components of E2E delay in network-based AR/VR

Table 2 shows approximate current values and projected values for each component of E2E delay, based on likely technology advances in hardware and software.

The current network transport latency is comprised of physical propagation delay and switching/forwarding delay at each network device.

1. The physical propagation delay: This is the delay caused by the speed limit of signal transmitting in physical media. Take the fiber as example, the optical transmit cannot exceed the light speed, or, 300km/ms in free space. But, light moving through the fiber optic core will travel slower than light through a vacuum because of the differences of the refractive index of light in free space and in the glass. In normal optical fiber, the light speed is about 200km/ms [Fiber-Light-Speed].

2. The switching/forwarding delay: This delay normally is much more than the physical propagation delay, which can vary from 200us to 200ms at each hop.

Latency	Current value (ms)	Projected value (ms)
T1	1	1
T2	11	2
T3	110 to 1000	5
T4	0.2 to 100	?
T5	5	5
T6	1	0.01
MTP	130 to 1118	13 + ?

MTP = T1+T2+T3+T4+T5+T6

Table 2: Current and projected latency in key stages in network based AR/VR

We can see that MTP latency is currently much greater than 20 ms.

If we project that the technology development and advance would bring down the latency in some areas, such as reducing the latency caused by GOP framing and streaming dramatically down to 5ms by using improved parallel hardware processing, and reducing display response time (refreshing latency) to 0.1 us by using OLED, etc; then the budget for the round trip network transport latency will be about 5 to 7 ms.

This budget will be consumed by propagation delay, switching delay and queuing delay. We can conclude

1. The physical distance between user and AR/VR server is limited and MUST be less than 1000km. So, the deployment of AR/VR server SHOULD be close to user as much as possible.

2. The total delay budget for network device will be low single digit, i.e. if the distance between user and AR/VR server is 600KM, then the accumulated maximum delay (round trip) allowed for all network devices is about 2 to 4ms. This is equivalent to 1 to 2ms delay in one direction for all network devices on the path.

A.2. Throughput Requirements

The Network bandwidth required for AR/VR is the actual TCP throughput required by application if the AR/VR stream is transported by TCP. It is another critical parameter for the quality of AR/VR application.

The AR/VR network bandwidth depends on the raw streaming data rate, or the bit rate for the video stream.

A.2.1. Average Throughput

The average network bandwidth for AR/VR is the average bit rate for AR/VR video.

For AR/VR video stream, there are many parameters that can impact the bit rate, such as display resolution, 2D or 3D, normal view or panorama view, the codec type for the video processing, the color space and sampling algorithm, the video pattern, etc.

Normally, the bit rate for 3D is approximately 1.5 times of 2D; and the bit rate for panorama view is about 4 times of normal view.

The latest codec process for high resolution video is H.246 and H.265. It has very high compression ratio.

The color space and sampling used in modern video streaming are YUV system [YUV] and chroma subsampling [Chroma].

YUV encodes a color image or video taking human perception into account, allowing reduced bandwidth for chrominance components, thereby typically enabling transmission errors or compression artifacts to be more efficiently masked by the human perception than using a "direct" RGB-representation.

Chroma subsampling is the practice of encoding images by implementing less resolution for chroma information than for luma information, taking advantage of the human visual system’s lower acuity for color differences than for luminance.

There are different sampling systems depends on the ratio of different samples for colors, such as Y’CrCb 4:1:1, Y’CrCb 4:2:0, Y’CrCb 4:2:2, Y’CrCb 4:4:4 and Y’CrCb 4:4:0. The most widely used sampling methods is Y’CrCb 4:2:0, this is often called YUV420 (note, the similar sampling for analog encoding is called Y’UV).

The video pattern, or motion rank, will also impact the stream bit rate. The video frames change more frequent, the less data compression will be obtained.

Compressed video stream consists of ordered successive group of pictures, or GOP [GOP]. There are three types of pictures (or frames) used in video compression, , such as H.264:

Intra code picture, or I-frames [GOP], Predictive coded picture, or P-frames [GOP] and Bipredictive coded picture, or B-frames [GOP].

An I-frame is in effect a fully specified picture, like a conventional static image file. P-frames and B-frames hold only part of the image information, so they need less space to store than an I-frame and thus improve video compression rates. A P-frame holds only the changes in the image from the previous frame. P-frames are also known as delta-frames. A B-frame saves even more space by using differences between the current frame and both the preceding and following frames to specify its content.

A typical video stream have a sequence of GOP with pattern, for example, IBBPBBPBBPBB, or, IBBBBPBBBBPBBBB.

The real bit rate also depends on the quality of the image user like to view. The Peak signal-to-noise ratio, or PSNR [PSNR] is to denote the quality of a image. The higher the PSNR, the better quality of the image, and the higher the bit rate.

Since human can only distinguish some level of image quality difference, it would be efficient to network if we could provide image with minimum PSNR that human eye perception cannot distinguish with image having higher PSNR. Unfortunately, this is still a research topic and there is no fixed minimum PSNR applies all people.

So, there is no exact formula for the bit rate, however, we can have experimental formula for the rough estimation of the bit rate for different parameters.

Formula (1) is from the H.264 Primer [H264_Primer]:

```
     Information rate = W * H * FPS * Rank * 0.07,    (1)

   where:
      W:    Number of pixels in horizontal direction
      H:    Number of pixels in vertical direction
      FPS:  Frames per second
      Rank: Motion rank, which can be:
            1: Low motion: video that has minimal movement
            2: Medium motion: video that has some degree of movement
            4: High motion: video that has a lot of movement and
               movement is unpredictable
```

The four formulae tagged (2) below are more generic and with more parameters for calculation of approximate information rates:

```
     Average information rate = T * W * H * S * d * FPS / Cv )
     I-frame information rate = T * W * H * S * d * FPS / Cj )
     Burst size = T * W * H * S * d / Cj                     ) (2)
     Burst time = 1/FPS                                      )

   where:
      T:   Type of video, 1 for 2D, 2 for 3D
      W:   Number of pixels in horizontal direction
      H:   Number of pixels in vertical direction
      S:   Scale factor, which can be:
            1   for YUV400
            1.5 for YUV420
            2   for YUV422
            3   for YUV444
      d:   Color depth bits
      FPS: Frames per second
      Cv:  Average compression ratio for video
      Cj:  Compression ratio for I-frame
```

Table 2 shows the bit rate calculated by the above formula 2 for different AR/VR levels.

It MUST be noted that in the Table 2:

1. There is no industry standard about the type of VR yet. The definition in the table is simply based on the 4K, 12K and 24K videos for 360x180 degree display. The Ultimate VR is roughly corresponding to the so called "Retina Display" which is about 60 PPD (Pix per degree) or 300 PPI (Pix per inch). However, there is argument about what is the limit of the human vision. J. Blackwell of the Optical Society of America has determined in 1946 that the resolution of the

human eye was actually closer to 0.35 arc minutes, which is more than 3 times of the Apple’s Retina Display (60 PPD).

2. The Mean and Peak Bit Rate illustrated in the table is calculated for a specific video with the acceptable perceptive PSNR, and with the typical compression ratio. It does not represent all type of videos. So, the compression ratio in the table is not universally applicable to all videos.

3. It MUST be aware that in the real use case, there are many schemes to reduce the video bit rate further in addition to the mandatory video compression. For example, only transmit the expected resolution for the video in the FOV in time, but transmit the video in other areas in slower speed, lower quality and lower resolution. All these technologies and their impact to the bandwidth are out of the scope of the document.

4. We assume the whole 360 degree video is transmitted to user site. The same video could be viewed by naked eye, or by HMD (without too much processing power). Thus, there is no difference to the network in bit rate, burst and burst time; The only difference is that using HMD can only view the video limited by its view angle. But if the HMD has its own video decoder, powerful processing and can directly communicate with the AR/VR content source, the network only needs to transport the data defined by HMD resolution which is only a small percentage of the whole 360 degree video. The corresponding data for mean/peak bit rate, burst size can be easily calculated by the formula (2). The last row "Infor Ratio of HMD/Whole video" denotes the ratio of Information amount (mean/peak bit rate and burst size) between HMD and the whole 360 degree video.

	Entry-level VR	Advanced VR	Ultimate VR
Type	4K 2D Video	12K 2D Video	24K 3D Video
Resolution W*H 360 degree video	3840*1920	11520*5760	23040*11520
HMD Resolution/ view angle	960*960/ 90	3840*3840/ 120	7680*7680/ 120
PPD (Pix per degree)	11	32	64
d (bit)	8	10	12
Cv	120	150	200(2D), 350(3D)
FPS	30	60	120
Mean Bit rate	22Mbps	398Mbps	2.87Gbps(2D) 3.28Gbps(3D)
Cj	20	30	20(2D), 30(3D)
Peak bit rate	132Mbps	1.9Gbps	28.7Gbps(2D) 38.2Gbps(3D)
Burst size	553K byte	4.15M Byte	29.9M Byte(2D) 39.8M Byte(3D)
Burst time	33ms	17ms	8ms
Infor Ratio of HMD/Whole Video	0.125	0.222	0.222

Table 2 Bit rate for different VR (use YUV420 and H.265)

A.2.2. Peak Throughput

The peak bandwidth for AR/VR is the peak bit rate for an AR/VR video. In this document, It is defined as the bit rate required to transport an I-frame, and the burst size is the size of I-frame, burst time is the time the I-frame must be transported from end to end based on FPS.

Similar to the Mean Bit rate, the calculation of Peak bit rate is purely theoretical and does not take any optimization into account.

There are two scenarios that a new I-frame will be generated and transported. One is when the AR/VR video display has dramatically changes that there is no similarity between two images; Another is when the FOV changes.

When AR/VR user is moving header or moving his eyeball to change Range of Interest, the FOV will be changed. FOV change may lead to the re-transmit of a new I-frame

Since there is no reference frame for the video compression, the I-frame can only be compressed by the infra-frame processing, or the compression for a static image like JPEG, and the compression ratio is much smaller than the inter-frame compression ratio.

It is estimated that the normal quality JPEG compression is about 20 to 30, This is only a fraction of the compression ratio for the normal video streaming.

In addition to the low compression issue, there is another problem involved. Due to the limit of MTP, the new I-frame must be rendered, grouped, transmitted and displayed in the delay budge for the network transport. This will cause the peak bit rate and burst size much bigger than the normal video streaming like IPTV.

The peak bit rate or the bit rate for I-frame, burst size and burst time are shown in the Formula 2. From the formula we can see the ratio of peak bit rate and the average bit rate is the ration of Cv/Cj. Since the Cv could be 100 to 200 for 2D, but the Cj is only about 20 to 30, so, the peak bit rate is about 10 times of average bit rate.

Authors' Addresses

Lin Han (editor)
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

Phone: +10 408 330 4613
Email: lin.han@huawei.com

Steve Appleby
BT
UK

Email: steve.appleby@bt.com

Kevin Smith
Vodafone
UK

Email: Kevin.Smith@vodafone.com