Guidelines For Video Delivery Over A Mobile Network

--For efficient video delivery--

Version 1.0

July 1st, 2014
NTT DOCOMO, Inc.

Contents

About trademarks

- All the names of the companies, products, and services that are mentioned in this document are trade names, trademarks, or registered trademarks of their respective owners.
- This document does not specifically indicate any copyrights, or trademarks or registered trademarks.

Revision history

Version	Section	Type	Revisions
1.0	—		First version

Chapter 1. Introduction

This document describes ways of providing video content over a mobile network aimed at providing comfortable video viewing experience for users. For comfortable video viewing experience, it is important to enable users to watch high-resolution video that runs smoothly without interruptions. However, the use of excessively high bitrates compared with the forwarding bandwidth that the network can provide may cause video images to skip or to be delayed, and this may result in deterioration in the service quality experienced by users. Therefore, it is important, especially in a mobile network, to maintain a suitable balance in the quality and size of images in video delivery.

While some measures are available to improve the quality of experience, the focus of this document is placed on those that can be used to optimize content bit rate and delivery bitrate. The first part of this document presents the fact that even for the same video, its data size varies depending on the compression measures used, showing some examples. The second part introduces some background information of a mobile network and its varying traffic volumes according to the time of day, and then moves on to discuss specific measures to control image quality based on the time of day. The third part focuses on HTTP adaptive streaming technology, which can provide smooth and seamless video streaming and efficiently utilize network resources in a mobile environment.

The use of measures introduced in this document will help achieve efficient video delivery. In addition to achieving comfortable video viewing experience to users, this will generate circumferential benefits through the reduction of their traffic volumes in the network. Efficient use of the finite resources in a mobile network will not only allow users to watch video comfortably but also enable others around them to use the network comfortably through suitable allocation of network resources. If your company is currently providing or planning to provide video content, this document will serve as a useful guide for achieving video delivery.

1.1. Target of this document

This document was written with video content providers in mind and, therefore, it presumes that the reader has a basic knowledge of video delivery.

1.2. Structure of this document

Chapter 2 of this document explains some unique characteristics of a mobile network. Chapter 3 describes schemes for image quality evaluation. Chapter 4 explains technological trends in video delivery. The background information provided in Chapters 2, 3, and 4 will help readers understand the effectiveness of specific measures introduced in Chapter 5. Those who are more interested in understanding specific measures and implementation methods can directly start reading Chapter 5 before the preceding Chapters 2, 3, and 4. Chapter 5 provides a systematic description of specific measures. Lastly, Chapter 6 presents some measures actually implemented by NTT DOCOMO as reference information. These examples are useful to have a better understanding of the details described in this document.

Please note that this document does not cover descriptions of server applications used for video delivery or terminal apps for video reception. This is due to an assumption that server applications and terminal apps that can use the technologies described in this document will be addressed by content providers engaged in video delivery.

1.3. References

[1] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. A. Joseph, A.Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," Proc. of ACM SIGCOMM'11, pp. 362-373, Aug.2011.
[2] ITU-T H.265 | ISO/IEC 23008-2 High Efficiency Video Coding
[3] ITU-T H.264 | ISO/IEC 14496-10 Advanced video coding for generic audiovisual services
[4] Jun Okamoto and Takanori Hayashi, “Latest trends in image media quality assessment technologies,” IEICE Fundamentals Review, Vol.6, No.4, pp.276-284, April 2013.(in Japanese)
[5] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” April 2008.
[6] ISO/IEC 23009-1 Information technology – Dynamic adaptive streaming over HTTP(DASH) – Part1: Media presentation description and segment formats

Chapter 2. Unique characteristics of mobile network services and actions taken by DOCOMO

2.1. Data volume in the mobile network by area and time period

Characteristics of a mobile network include varying peak hours depending on the characteristics of areas* (see Figure 2-1 below). As of December 2013, the level of congestion in the DOCOMO network peaks from 10 p.m. to 11 p.m. This is based on the total volume of data generated by NTT DOCOMO's Xi (LTE) service subscribers going through the sp mode center. A closer analysis of smaller areas reveals that the peak time varies in different areas. The peak time is from 12 noon to 1 p.m. in an area covering a business district; from 7 a.m. to 8 a.m. and 6 p.m. to 7 p.m. at an urban terminal station area; and from 12 midnight to 1 a.m. in a residential area. This shows that each area experiences congestion in a different time zone. A failure to use an optimal bitrate for video delivery in respective areas and time zones will increase the risk of causing more intermissions and slowdowns in video. This makes users feel stress during video viewing (lowering the quality of experience for users). Therefore, in order to improve the quality of user experience, it is important to dynamically change the delivery bitrate to optimize video delivery in accordance with the degree of congestion in the network, using measures such as those hereinafter described.

*The "area" mentioned here is defined as an area covered by one or more base stations.

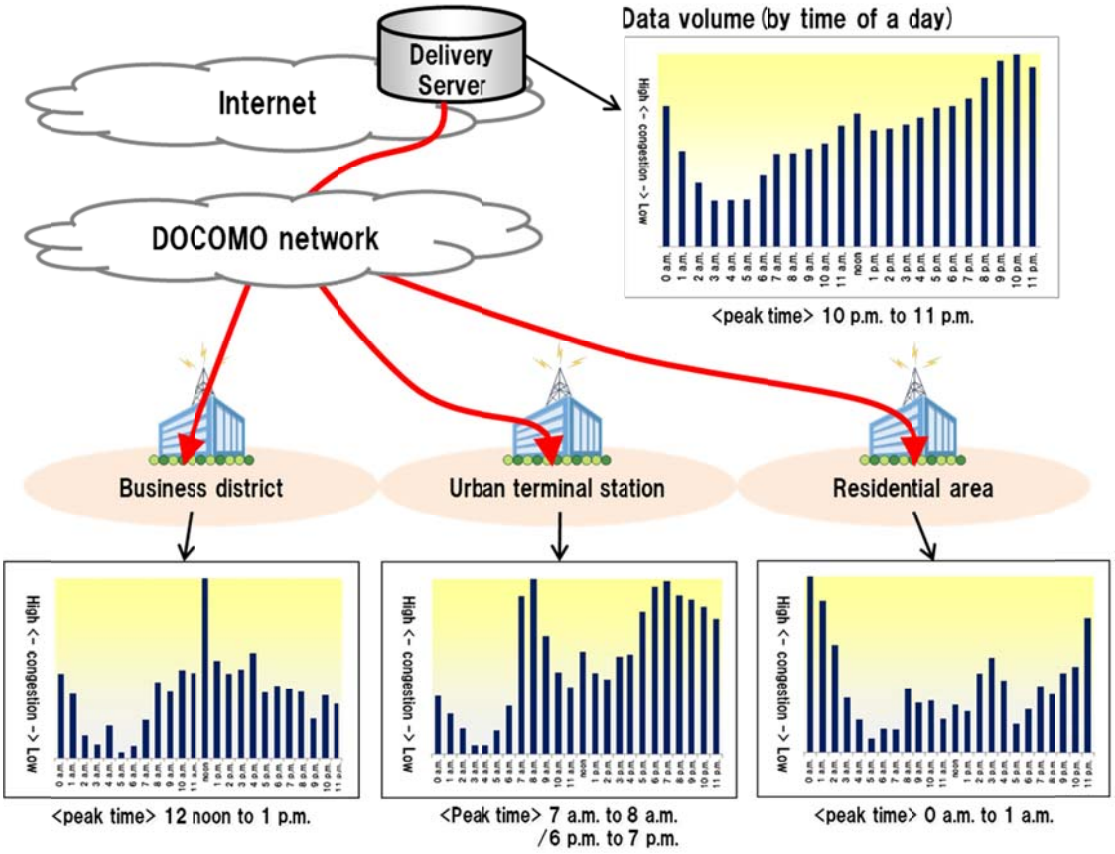


Figure 2-1 Changes in the data volume in a mobile network by area and time

2.2. Limit of transmission speeds up to 128 kbps

For NTT DOCOMO's Xi (LTE) service, if the volume of data used in the current month exceeds a certain threshold, a 128-kbps cap is imposed on the user's upload/download speed for the rest of the month. This limitation is one of the measures designed from the standpoint of ensuring fair network usage among users. The data volumes that trigger such limitation are shown in Tables 2-1 and 2-2. The use of an excessively high bitrate in video delivery therefore may invite the risk of putting the user at a disadvantage.

Table 2-1 Summary of major Xi Packet Flat-rate services (currently available)

Xi Packet Flat-rate Services	Max data volume for a normal transmission speed	Requirement for removing the 128kbps cap
Xi Pake-hodai Light	3GB/month	Payment of extra charges for every 2GB data usage
Xi Pake-hodai Flat	7GB/month	
Xi Pake-hodai for iPhone		

*For details of Xi Packet Flat-rate Servcies, please refer to NTT DOCOMO's web page (URL: https://www.nttdocomo.co.jp/charge/packet/index.html)

Table 2-1 Summary of major Packet Packs in new billing plans (available from June 1, 2014)

Packet Pack	Max data volume for a normal transmission speed	Requirement for removing the 128kbps cap
Raku-Raku Pack	200MB/month	Payment of extra charges for every 1GB data usage
Data S Pack	2GB/month	
Data M Pack	5GB/month	
Share Pack 10	10GB/month (per group)	
Share Pack 15	15GB/ month (per group)	
Share Pack 20	20GB/ month (per group)	
Share Pack 30	30GB/ month (per group)	

*For details of the new billing plans "Kake-hodai & Pake-aeru," which will help families save money, please refer to NTT DOCOMO's web page. (URL:https://www.nttdocomo.co.jp/charge/new_plan/index.html?icid=CRP_TOP_main PR_new_plan#charge)

2.3. NTT DOCOMO's actions to enhance its mobile network

NTT DOCOMO is enhancing its mobile network by efficiently using four spectrum bands (quad-band LTE) and installing more LTE base stations to support growing data volume.

In addition to the actions above taken by NTT DOCOMO, content providers can also play a role in making the users' video watching experience much more comfortable by using measures that will be described in Chapter 5 for video delivery. For actions taken by NTT DOCOMO to enhance its mobile network, please refer to NTT DOCOMO's web page (URL: https://www.nttdocomo.co.jp/xi/index.html)

2.3.1. Quad-band LTE (efficient use of 4 spectrum bands)

NTT DOCOMO's LTE "Xi" service efficiently utilizes the four spectrum bands - 1.7GHz, 1.5GHz, 800MHz, and 2GHz- in order to further enhance its mobile network.

The 2GHz and 800MHz bands have been used to provide "coverage" and the 1.5GHz band to provide "speed." In September 2013, the 1.7 GHz band was added to NTT DOCOMO's 3 operation bands mentioned above to launch its "quad-band LTE" operation.

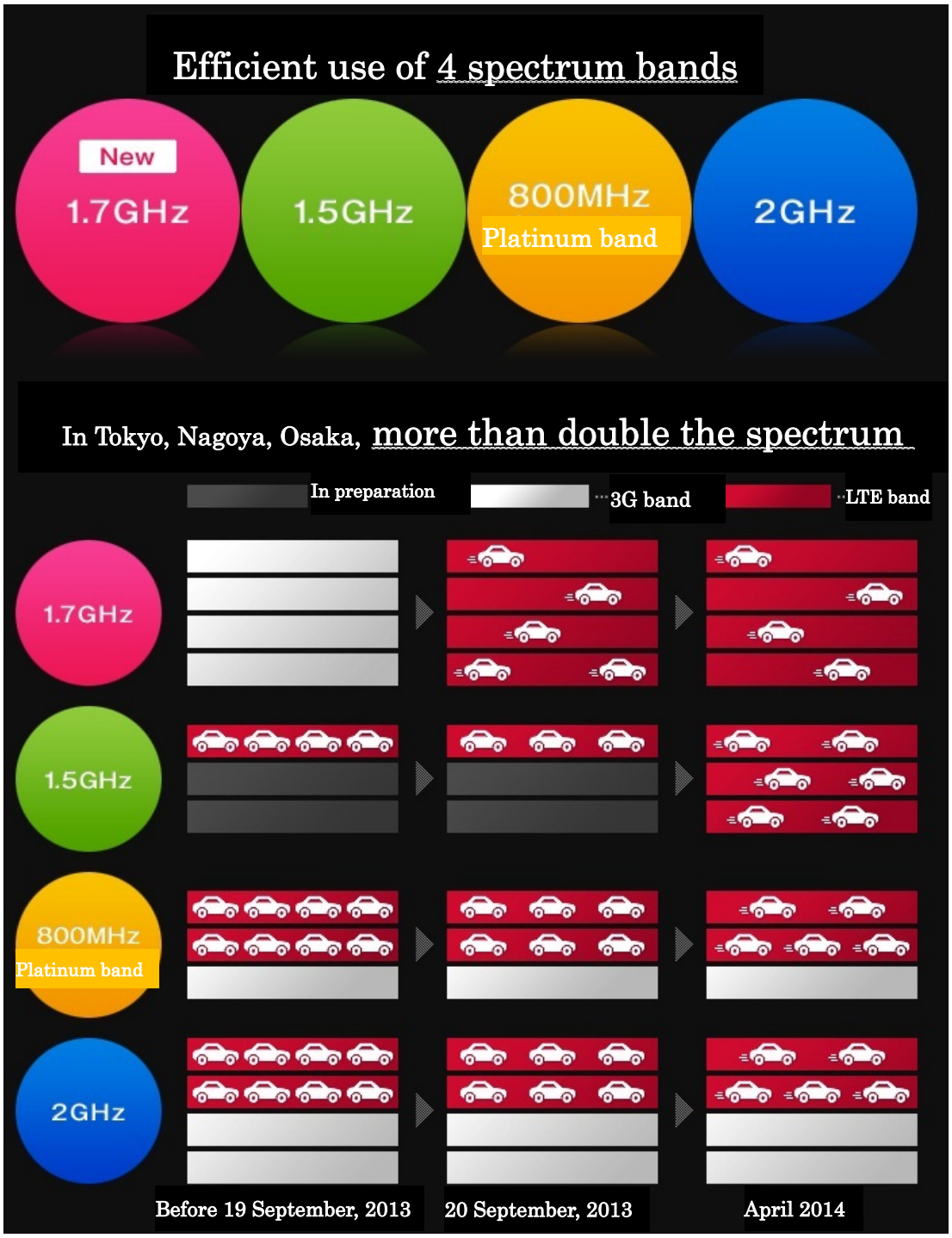


Figure 2-2 Image of quad-band LTE

2.3.2. Installation of additional LTE base stations

NTT DOCOMO is installing more LTE base stations to expand its LTE "Xi" areas so that connections will be available anytime, anywhere. The number of its LTE base stations will be increased 1.7 times over the fiscal year 2014 (from approx. 55,300 at the end of FY 2013 to approx. 95,300 at the end of FY 2014) and the number of 100Mpbs base stations more than 10 times (from approx. 3,500 to 40,000 during the same period as above) for further enhancement of its LTE network.

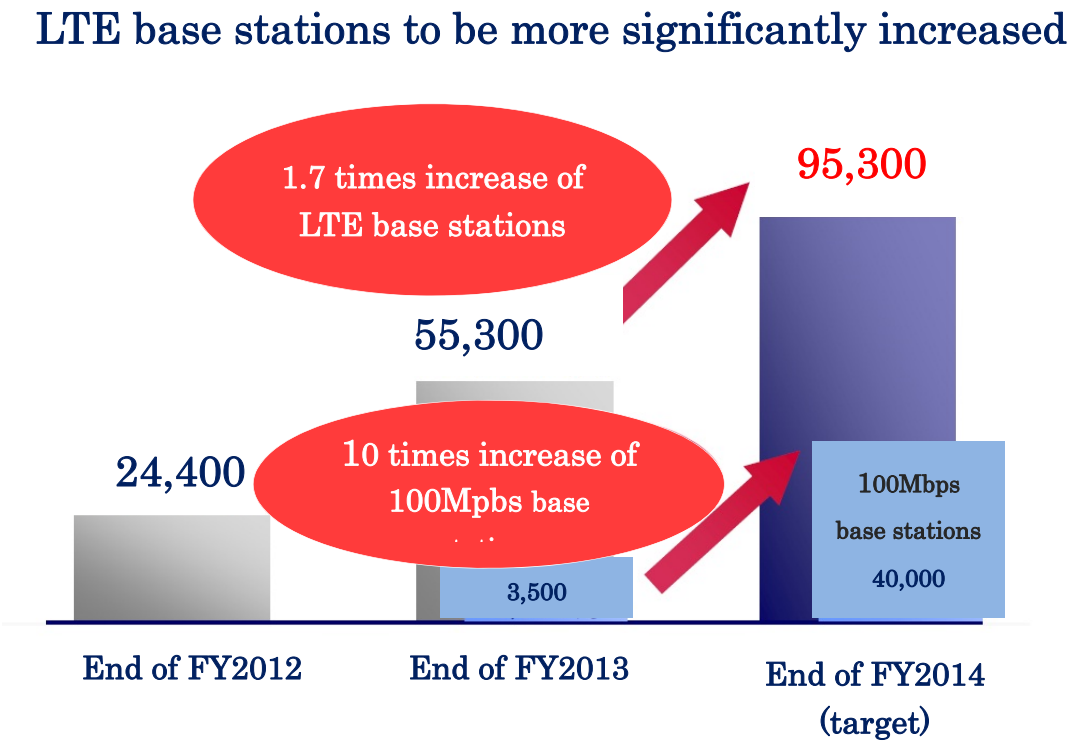


Figure 2-3 LTE Xi base station installation plan

Chapter 3. Image quality experienced by customers using video delivery service

3.1. Factors affecting the image quality of video delivery service

This section describes some factors that affect the quality of images experienced by users (Quality of Experience) in video delivery service.

Figure 3-1 shows elements that constitute video delivery service using a mobile network as well as major factors that affect the video image quality. Video data is encoded into a format suitable for transmission over the network and stored in the delivery server of a content provider. NTT DOCOMO's network receives video data over the Internet, etc. and forwards it via a radio link such as an LTE network to smartphones, tablets, and other video viewing terminals. The terminal that receives the video data decodes the data into images that are to be displayed on its screen.

The image quality of video delivery service is affected by different elements that constitute video delivery service. Major factors that affect the image quality of video delivery service are generally classified as follows:

(1) Image encoding conditions;
(2) Throughput characteristics of transmission of encoded data over a mobile network;
(3) Performance of the video viewing terminal.

First, (1) Image encoding conditions for video delivery include the following.

・Coding scheme
・Encoding bitrate
When the bitrate is set low, it results in lack of sharpness (blurring) or mosaic-like distortions (, which is called "block-noise") in images.
・Image resolution
When the image resolution is low, it reduces the definition of images.
・Frame rate
When the frame rate is low, it reduces the smoothness of motions, resulting in flickering and jerky images.

Second, when (2) the throughput of the mobile network is too low, the download of video data cannot catch up with the playback of video on the terminal. As a result, the playback is halted until a certain amount of video data are stored again in the terminal

(, which is called rebuffering). Such throughput degradation can be caused by packet loss, packet transmission delay, and delay variation that occur in the network.

Furthermore, regarding (3) the performance of the video viewing terminal, the image quality achieved on the terminal differs depending on their decoding and display capabilities. Compared with smartphones, tablet terminals have larger screens. This makes image deterioration on tablet terminals more easily perceived even when both terminals are showing video with the same image quality,.

The factors affecting image quality are not mutually independent but related. For instance, a higher encoding bitrate generally improves image quality. However, if an unnecessarily high bitrate is used for video data transfer, it may cause network congestion (throughput degradation). Achieving efficient video delivery considering the level of congestion in a mobile network requires a deep understanding of such quality factors.

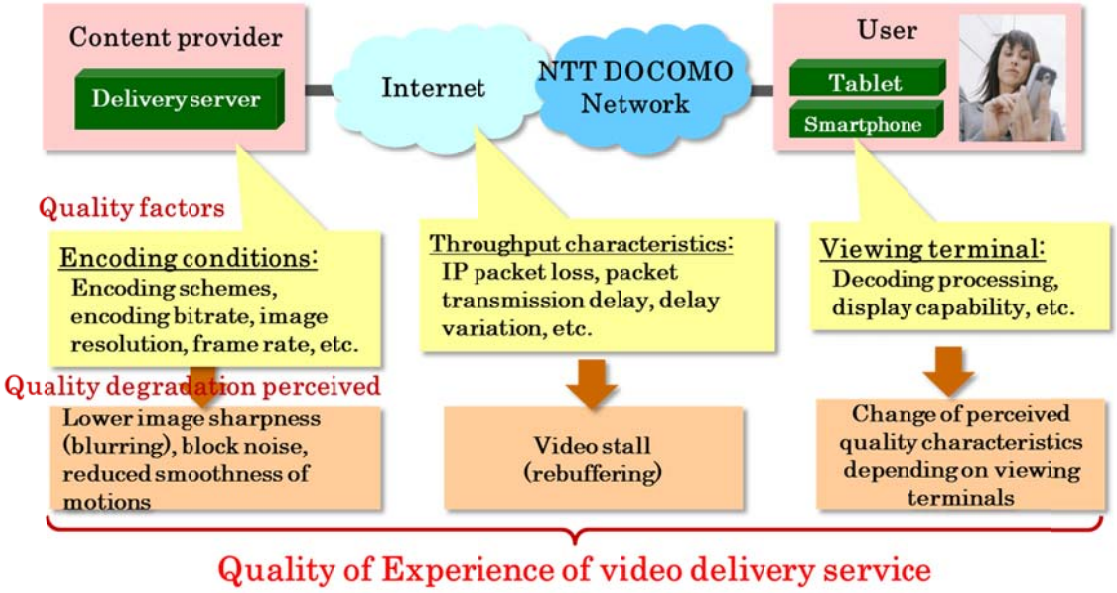


Figure 3-1 Major factors affecting the elements and image quality of video delivery service using a mobile network

3.2. Relationship between encoding conditions and image quality

Understanding of the relationship between encoding conditions and image quality is first and foremost important to deliver video to users with optimized image quality in accordance with the level of congestion in a mobile network. Please note that encoding conditions hereinafter described include the coding schemes, encoding bitrates, image resolutions, and frame rates to be adopted. In this section, the relationship between

encoding bitrate and image quality and the relationship between image resolution/frame rate and image quality are discussed.

3.2.1. Relationship between encoding bitrate and image quality

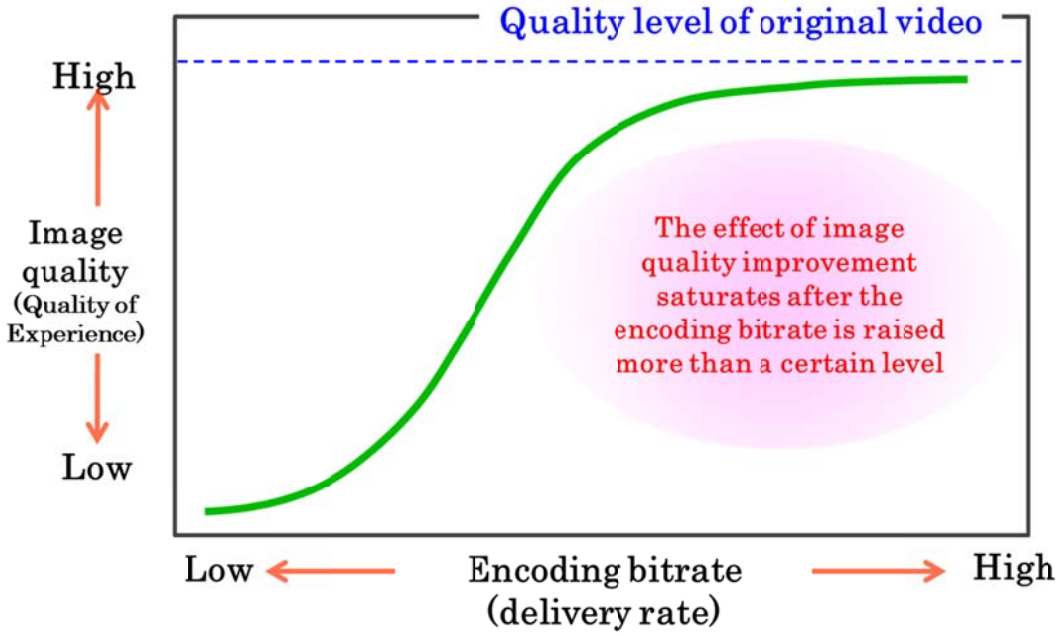


Figure 3-2 shows a general relationship between encoding bitrate and image quality. The horizontal axis shows the encoding bitrate while the vertical axis shows the image quality. The higher the encoding bitrate is set, the closer the image quality gets to that of the original video (the video before encoding). However, the image quality comparable to that of the original video can be obtained when a certain level of encoding bitrate are secured. On the other hand, the effect of image quality improvement becomes more marginal when the encoding bitrate is set too high beyond necessity

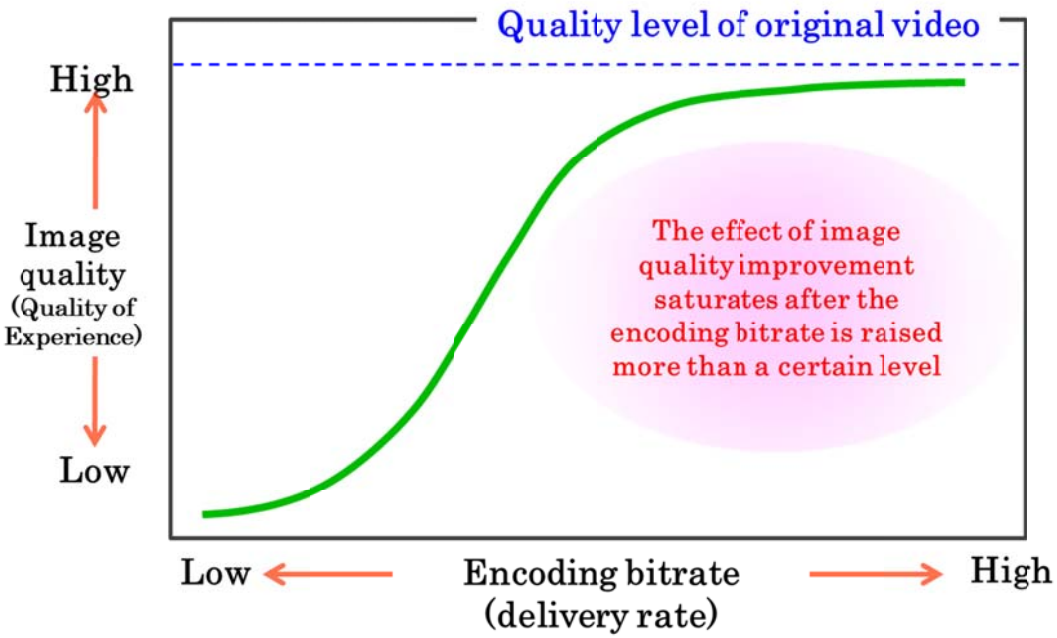
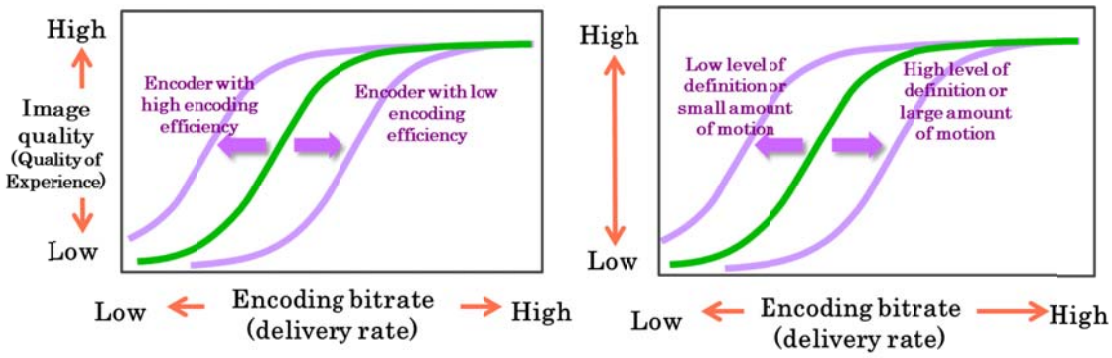


Figure 3-2 Relationship between encoding bitrate and image quality

The value of encoding bitrate that gives sufficient image quality varies among coding schemes. Even when the same coding scheme is applied, the image quality is different[1] depending on encoder implementation (difference in the encoding efficiency). And even when the same encoder is used, the image quality is different depending on the level of definition and amount of motion of the video. In general, video with a lower level of definition and a smaller amount of motion can achieve high image quality at a lower bitrate. On the other hand, video with a higher level of definition and a larger amount of motion requires a higher encoding bitrate to achieve high image quality. These relationships are summarized in Figure 3-3.



[1]This is because the specifications (standards) of video coding schemes only target decoding and not encoding, trying to leave encoder manufacturers/software providers room for image quality improvement.

Figure 3-3 Effects of implementation of coding schemes and characteristics of video on the relationship between encoding bitrate and image quality

3.2.2. Relationship between image resolution/frame rate and image quality

An increasing number of content providers are handling high-resolution images such as those for high-definition television (HDTV) in order to provide users with high-resolution video delivery service. The current mobile network environment, however, is not always fully prepared to ensure the bandwidth sufficient for delivery of HDTV-class video. This makes necessary to select a suitable image resolution and frame rate in order to achieve higher video quality with a smaller bandwidth. Coding with a low image resolution can yield high quality image in a low bandwidth. This relationship is shown in Figure 3-4. In an example illustrated in Figure 3-4, the encoder using the bitrates B_1, B_2, and B_3 achieves their optimal video image quality for the image resolutions of 320×240, 640×480, and 1280×720, respectively. When an encoding bitrate lower than a level of 100 to 200kbps is used, reducing the frame rate is effective from the standpoint of maintaining the image quality. However, reducing the frame rate in fast-motion video such as of a sport scene will make the flickering and jerky parts more noticeable. Degradation in the quality of experience can be minimized for content like video telephony even at a low frame rate. However, for content of ordinary video delivery service, a frame rate of approximately 15 frames per second is at least necessary.

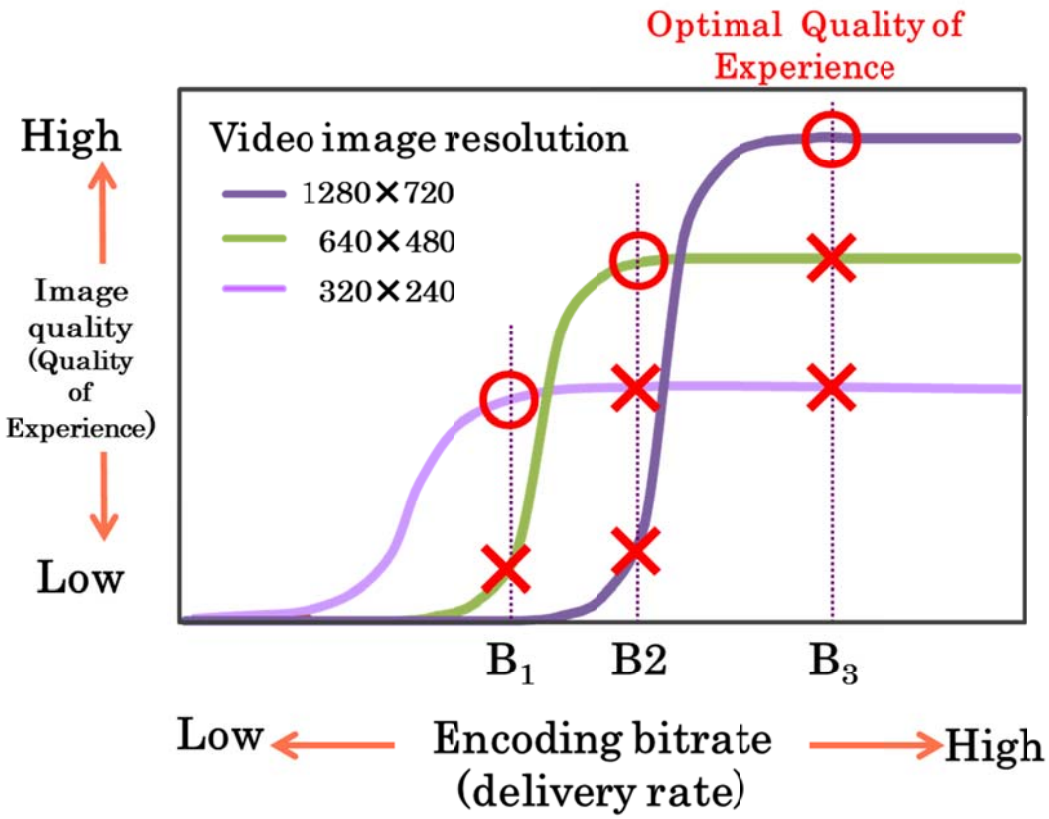


Figure 3-4 Relationship between image resolution and image quality

3.3. Degradation in quality of experience in a congested mobile network

This section describes characteristics of degradation in the quality of experience that occurs when a mobile network is congested. As was shown in

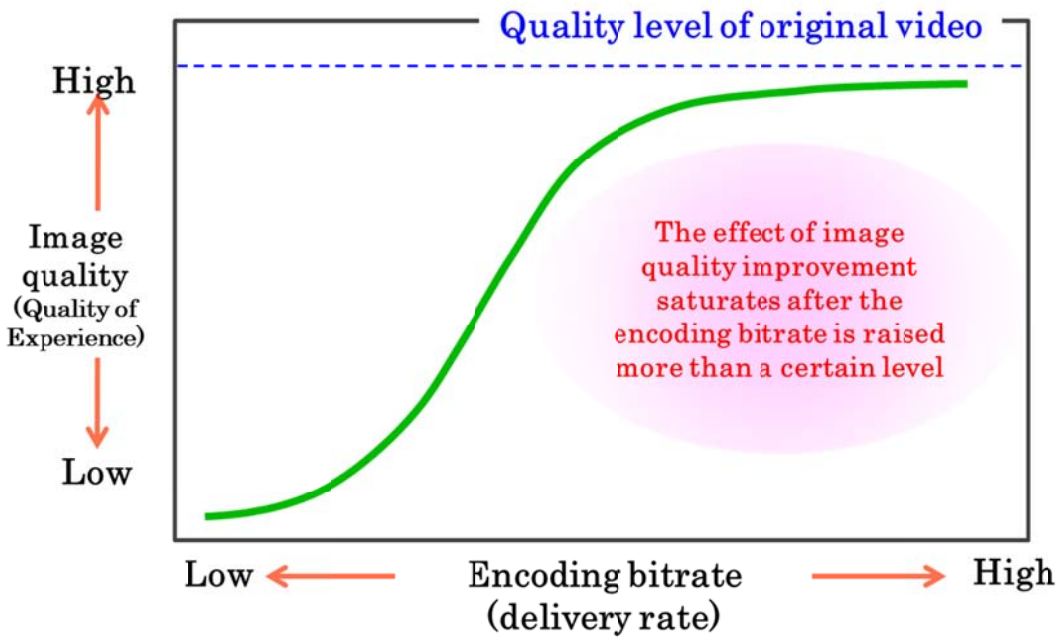


Figure 3-2 above, when a sufficient throughput is secured in a mobile network, the image quality gradually improves as the encoding bitrate is raised until such improvement comes to saturation at a certain point. However, even when higher image quality is achieved with a higher encoding bitrate, video playback may frequently stall in a congested mobile network due to insufficient throughput. This causes a drastic deterioration in the user's quality of experience. Figure 3-5 demonstrates an example in which degradation in the quality of experience occurs at the throughputs of 300 and 500 kbps. In both cases, the use of encoding bitrate larger than the respective throughput causes video playback to stall, which lowers the quality of experience. Such video stall not only degrades the quality of experience but also greatly affects the length of video watching time (video viewing behavior) of users according to some reports (reference [1], etc.). Therefore, it is necessary to prevent video stall in order to maintain and improve the quality of video images.

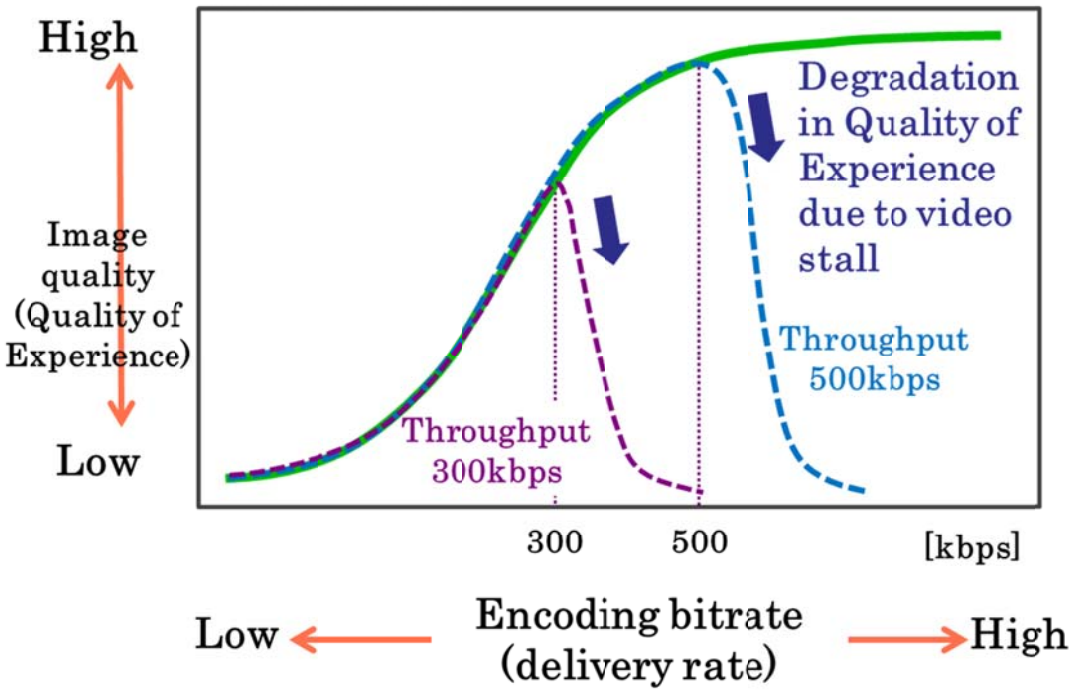


Figure 3-5 Degradation in quality of experience due to video stall

Chapter 4. Technological trends in video delivery

4.1. High-compression codec technologies

The latest video compression standard HEVC or officially called "ITU-T H.265 | ISO/IEC 23008-2 High Efficiency Video Coding[2]" was approved by the ITU in January 2013. HEVC is an all-purpose standard designed to address a wide range of applications from video on small screens of mobile devices to high-definition video with HD resolution of 4K, 8K or higher. The HEVC standard was developed to double the compression rate compared with the conventional H.264/AVC standard[3].

First, this section describes the terms "Profile," "Level," and "Tier," which denote the playback capability of an H.265/HEVC decoder.

The current H.265/HEVC standard defines the following three Profiles:

-Main Profile;

-Main 10 Profile;

-Main Still Picture Profile.

Both of Main Profile and Main 10 Profile are profiles for video compression, using different bit depths to express one pixel. Main Profile only supports video with a depth of 8 bits per pixel while Main 10 Profile also supports 10 bits per pixel in addition to 8 bits per pixel. Except for bit depth coding, the both Profiles use common technical elements for coding. Both of them only support the chroma subsampling format of 4:2:0 (in which the amount of information on color difference is a quarter of that of brightness), which is for consumer applications. An extension of the H.265/HEVC standard called RExt is being standardized to support formats with a large color difference, such as 4:2:2 and 4:4:4, which are mainly used for material transmission in broadcasting stations. HEVC RExt also aims to support video with a bit depth higher than 10bits per pixel. Thus, HEVC RExt is a standard mainly targeted at compression and transmission of high-definition video, such as digital cinema and material transmission. As for interlaced video, Main Profile and Main 10 Profile can handle this type of video . However they haven't introduced any coding tools specialized for interlaced video and currently only use flags to mark it in the stream. In actual processing, interlaced video is encoded as progressive video with its top and bottom fields separated.

Next part discusses the Level and Tier. There are 13 Levels from 1 to 6.2, each of which mainly specifies the maximum resolution and frame rate. The Tier, on the other hand, is

divided into two, the main Tier and the high Tier, each of which specifies the maximum bitrate at each level. H.264/AVC does not have a concept of Tier but specifies everything including the maximum resolution, frame rate, and bitrate, using the Level only. However, this makes it difficult to address a case where only the bitrate should be restricted. H.265/HEVC, which has newly introduced the concept of Tier, is able to specify the restriction on the resolution, frame rate, and bitrate, separately. Table 4-1 below shows a list of major resolutions, and Levels and Tiers.

Table 4-1 List of main resolutions, Levels, and Tiers

Main Levels	Examples of maximum resolution and frame rate	Maximum video bitrate [kbps]	
		Main Tier	High Tier
1	QCIF 15fps	128	–
2	CIF 30fps	1,500	–
3	QHD(960x540) 30fps	6,000	–
4	2Kx1080(2048x1080) 30fps	12,000	30,000
5	4096x2160 30fps	25,000	100,000
6	8192x4096 30fps	60,000	240,000
6.2	8192x4096 120fps	240,000	800,000

Listed below are specific examples of coding technologies used in H.265/HEVC.

- Block partitioning
- Inter and intra prediction
- Orthogonal transform
- Variable-length coding
- In-loop filter

As is the case with major conventional coding standards, H.265/HEVC adopts a compression scheme based on a combination of motion compensating prediction and orthogonal transform. H.265/HEVC doubles the compression performance of H.264/AVC by accumulating improvements in individual coding tools. However, H.265/HEVC has no compatibility with conventional standards. Furthermore, its complexity gradually adds more processing load, especially at the time of compression. To cope with this, researchers and developers are working on to develop an H.265/HEVC encoder capable of high-speed operation.

As shown in Figure 4-1, a hierarchical block portioning structure (Coding Tree Unit; CTU) is introduced to H.265/HEVC to enable efficient coding of video images in widely varying levels of resolution. Each block in the CTU is further split into subdivisions called PUs (Prediction Units). By combining all these blocks, a wide variety of blocks

can be used for coding, ranging from the minimum size of 4x4 to the maximum size of 64x64.

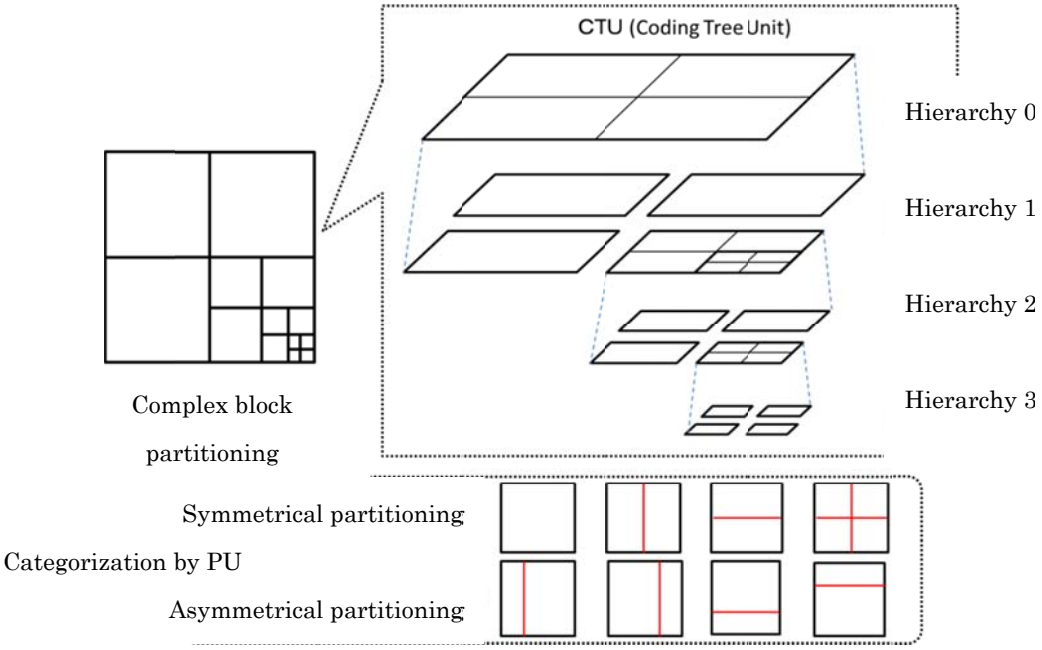


Figure 4-1 Block partitioning method of H.265/HEVC

The introduction of this method has enabled H.265/HEVC to perform complex block partitioning and optimized coding according to image characteristics in each block, leading to compression performance improvement.

As for intra prediction, the number of prediction modes available for H.265/HEVC has drastically increased to 35 from 9 for H.264/AVC. This allows more precise prediction in combination with the blocks of different sizes described above. Inter prediction (motion prediction), which has also increased block size options, has introduced some innovative measures. They include the use of a newly created merge mode and multiple prediction vector candidates. Those measures have contributed to achieving higher compression rates.

A hierarchical structure has also been introduced to orthogonal transform. The ability to combine four sizes of orthogonal transform from 4x4 to 32x32 has greatly improved compression efficiency.

As for variable-length coding, the only scheme available is the Context-Adaptive Binary Arithmetic Coding (CABAC), which has high compression performance. H.264/AVC has used the Context-based Adaptive Variable Length Coding (CAVLC), a lightweight scheme, in spite of its low compression efficiency. On the other hand, this option has

been excluded and is not available for H.265/HEVC in consideration of the progress of device performance in recent years.

For an in-loop filter, the deblocking filter for block noise reduction and the Sample Adaptive Offset (SAO) for ringing and pixel value shift control have been introduced. These have contributed to achieving a reduction of coding distortion.

Table 4-2 provides a comparative summary of coding tools between H.265/HEVC and H.264/AVC.

Table 4-2 Comparative summary of coding tools of H.264/AVC and H.265/HEVC

Coding tool	H.264/AVC (High Profile)	H.265/HEVC
Intra prediction	4x4 block: 9 modes 8x8 block: 9 modes 16x16 block: 4 modes	35 modes for 4x4 to 64x64 block sizes
Inter prediction (motion prediction)	4x4 to 16x16 blocks Quarter-pel accuracy search	8x4/4x8 to 64x64 blocks Quarter-pel accuracy search
Orthogonal transform	4x4 or 8x8 Integer accuracy DCT	4x4 to 32x32 (4 types) Integer accuracy DCT
Variable coding	2 types: ・Context-based Adaptive Variable Length Coding (CAVLC) ・Context-Adaptive Binary Arithmetic Coding (CABAC)	1 type: ・Context-Adaptive Binary Arithmetic Coding (CABAC)
In-loop filter	1 type: ・Deblocking filter	2 types: ・Deblocking filter ・Sample Adaptive Offset (SAO)

4.2. Video streaming technologies

Video streaming technologies are available for delivery of video content over a network . Video streaming technologies are mainly divided into two groups depending on the protocols they use: HTTP and non-HTTP (RTSP：Real-Time Streaming Protocol, etc.). Video delivery services today use HTTP-based streaming technologies. This is because

the HTTP protocol has excellent conformity with standard web servers and existing server systems in that it can make use of HTTP cache. HTTP is also highly compatible with user systems in that it can transparently go through firewalls and NATs. Most well-known HTTP-based streaming technologies are HTTP streaming and HTTP adaptive streaming.

In video delivery, the quality of experience of users is affected by the quality of video (resolution, frame rate, etc.) and the smoothness of playback. Raising the quality of video increases the amount of data (bitrate) flowing over the network. The more the video bitrate exceeds the transfer speed, the more frequently video playback will be interrupted. In this manner, video quality and playback smoothness are in the relationship of trade-off. Especially in a mobile environment, characterized by unstable and highly fluctuating transfer speed, our challenge is how to ensure playback smoothness. HTTP adaptive streaming is aimed at mitigating such trade-off.

4.2.1. HTTP streaming (pseudo streaming)

This is the most widely utilized streaming technology in video delivery service. In HTTP streaming, when the user starts video playback, it triggers the download of a selected video file from the web server. A video playback application (web browser or media browser) performs sequential playback of the buffered video data in parallel with the download of the video file.

The bitrate of the downloaded video data is fixed from the start to the end of its playback. When the bitrate of the specified video data is high, it takes a long time until the start of playback due to a large amount of data to be buffered. Although the amount of data in the buffer is reduced as the playback proceeds, the playback will be interrupted if the buffer is exhausted. To restart the playback, it is necessary to obtain buffered data once again. If the user cancels the playback at this point, the data not yet played back will be discarded. This means that the resources of the server and network used for the download will be wasted.

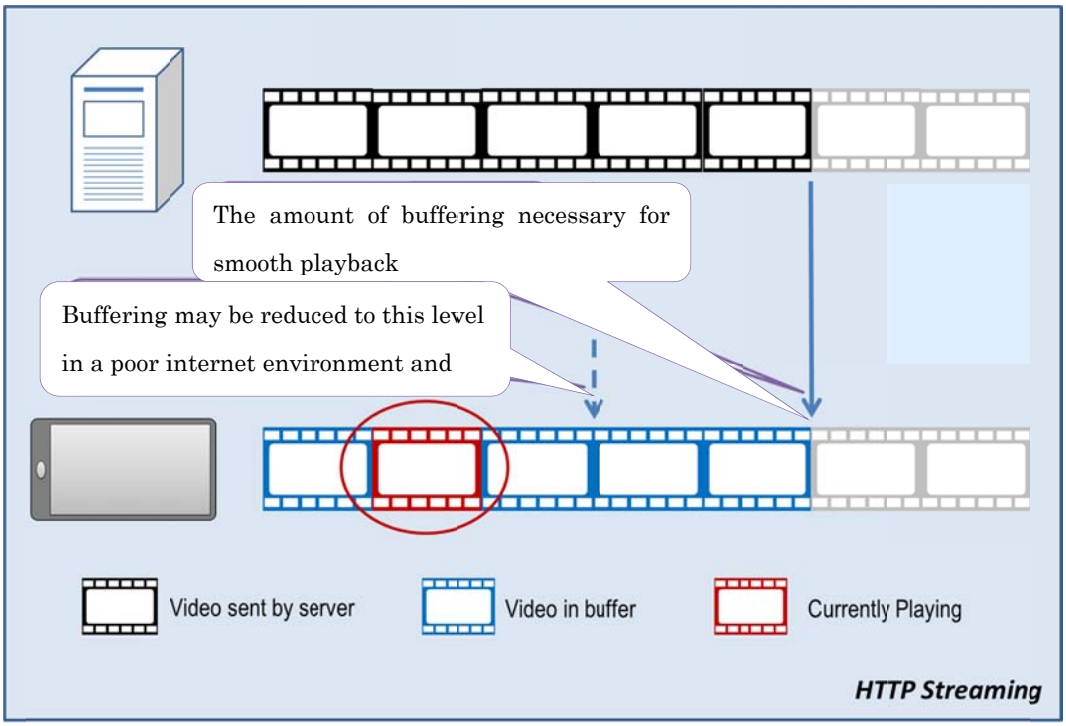


Figure 4-2 HTTP streaming

4.2.2. HTTP adaptive streaming

To deliver high-quality video to a variety of devices (PCs, TVs, smartphones, etc.) using various network environments (Wi-Fi, 3G/LTE, fixed networks, etc), a streaming technology that enables switching of bitrates according to specific network environment has been developed. The technology is called HTTP adaptive streaming.

Even in a mobile network, which features ever-changing network conditions, this streaming technology can achieve almost uninterrupted, smooth video streaming by delivering video data with a bitrate suitable to the network environment. This technology can reduce the gap between transfer speeds and video data bitrates. This makes it possible to reduce the amount of buffer and thus shorten the time required before the start of playback.

Furthermore, in this technology, the server need not send an entire video file at one time, but send a part much data of the file as requested by the device receiving the video. This reduces the amount of data discarded even when the user cancels the playback. Therefore, it can effectively reduce wasted of the server and network.

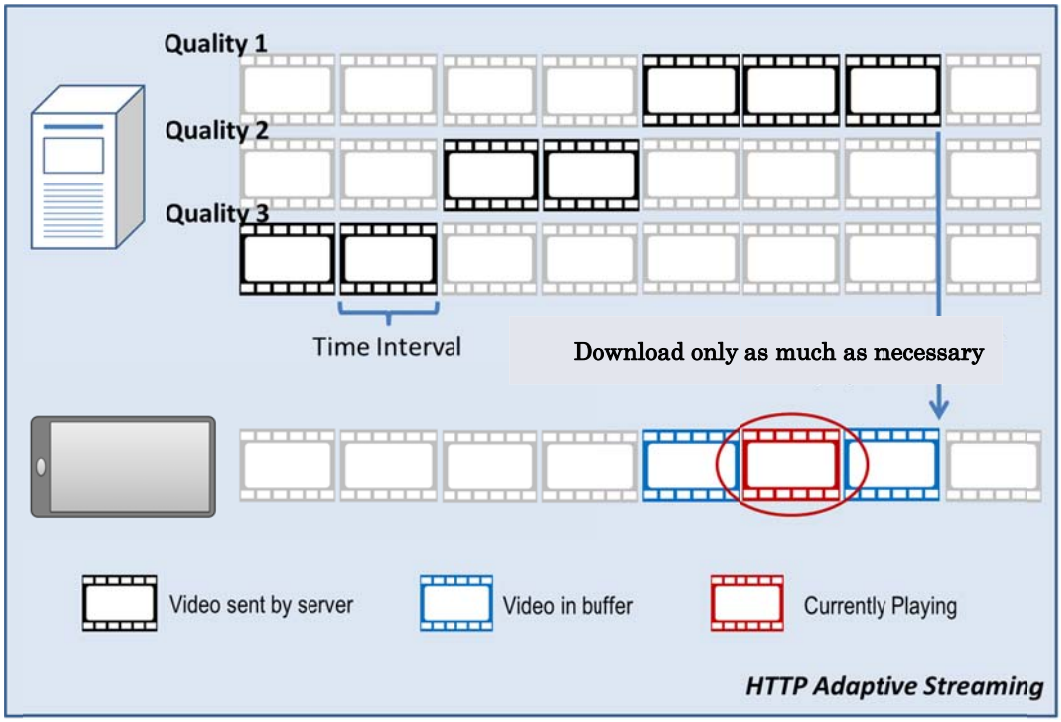


Figure 4-3 HTTP adaptive streaming

Chapter 5. DOCOMO-Suggested measures for efficient network utilization

This chapter introduces some recommended measures that should be considered for adoption to achieve efficient use of a mobile network for video delivery.

5.1. Optimization of content bitrate (image quality)

5.1.1. Optimization for H.264/AVC

Table 5-1 shows a general relationship between the video image quality and the DOCOMO-suggested image resolution and encoding bitrate for the H.264/AVC coding technology. The "image quality" in the table means the quality of images experienced by the user (quality of experience) who watches the video. The image quality in the table is quantified based on the subjective perception (subjective quality) of humans users, such as "clear image," "smooth video," and "image quality degradation perceptible but not annoying."[2] Table 5-1 shows a result of a case using Main Profile as the profile, a frame rate of 30fps, and a random access interval of approximately one second (32 frames). As described earlier, when the encoding bitrate becomes low, the use of low image resolution helps reduce image quality degradation. The table shows a sample case in which the image resolutions are changed corresponding to the encoding bitrates.

Table 5-1 Relationship between image quality and DOCOMO-suggested image resolution/encoding bitrate (in case of H.264/AVC)

Coding scheme: H.264/AVC	Image quality		
	Low	Middle	High
Suggested image resolution	360p-equivalent (640×360)	360p-equivalent (640×360)	720p-equivalent (1280×720)
Suggested video encoding bitrate	S: approx. 350kbps T/P: approx. 400kbps	S: approx.700kbps T/P: approx. 800kbps	S: approx. 1.4Mbps T/P: approx. 1.6Mbps

[2]The image quality was determined based on the results of evaluation conducted using the ACR Method (Absolute Category Rating Method), which is an evaluation method adopted as an international standard. Specifically, the quality of images of different types of content were evaluated. The suggested conditions for High, Middle, and Low image quality levels are derived from the content which showed average quality characteristics and evaluated to have ratings of 4, 3, and 2 respectively on the MOS (Mean Opinion Score) scale. For evaluation methods quantifying subjective image quality, please refer to the references [4] and [5], etc.

* S stands for Smartphones: evaluation results from video viewing on 4.7-inch smartphones.
  T/P stands for Tablets and PCs: evaluation results from video viewing on 15.5-inch notebook PCs.
* The values of the suggested video encoding bitrates are subject to change.

5.1.2. Optimization for H.265/HEVC

An effective measure to achieve high image quality at a lower encoding bitrate than that of H.264/AVC is to use a coding scheme with a high compression capability. The latest H.265/HEVC coding scheme discussed in Chapter 4 is said to almost double the compression efficiency compared to H.264/AVC. Therefore, H.265/HEVC is expected to be widely utilized in the mobile video delivery market in the future. Table 5-2 shows a general relationship between image quality and suggested image resolutions and encoding bitrates for the current H.265/HEVC coding technology. The table shows a result of study of a case that uses Main Profile as the profile, a frame rate of 30fps, and a random access interval of approximately one second (32 frames). Future improvements in H.265/HEVC implementation technology are expected to achieve even higher compression levels.

Table 5-2 Relationship between image quality and DOCOMO-Suggested image resolution/encoding bitrate (in case of H.265)

Coding scheme: H.265/HEVC	Image quality		
	Low	Middle	High
Suggested image resolution	360p-equivalent (640×360)	360p-equivalent (640×360)	720p-equivalent (1280×720)
Suggested video encoding bitrate	S: approx. 200kbps T/P: approx. 200kbps	S: approx.350kbps T/P: approx. 400kbps	S: approx. 700kbps T/P: approx. 800kbps

* S stands for Smartphones: evaluation results from video viewing on 4.7-inch smartphones.
  T/P stands for Tablets and PCs: evaluation results from video viewing on 15.5-inch notebook PCs.
* The values of the suggested video encoding bitrates are subject to change.
* In H.265/HEVC, when a high image resolution is used, coding may take a long time.

5.2. Optimization of delivery bitrate (Quality of Experience)

5.2.1. Video streaming technologies suitable for a mobile environment

Table 5-3 shows the status of support on Android and iOS platforms for the streaming technologies discussed in Section 4.2. HTTP Adaptive Streaming has multiple technologies. Although they share the same basic concept of switching content files according to the state of network, they adopt different file formats. MPEG-DASH (Dynamic Adaptive Streaming over HTTP)[6] is an ISO/IEC standard and is officially supported on Android4.4 and higher. HLS (HTTP Live Streaming) is a proprietary

technology of Apple Inc., although it is supported on both Android and iOS platforms. These two HTTP Adaptive Streaming technologies and HTTP streaming are the three technology options for video delivery for mobile devices. (As stated in Section 4.2, in light of a mobile environment, the adoption of HTTP Adaptive Streaming (MPEG-DASH or HLS) is expected to improve user experience.)

This section describes DOCOMO-suggested operation methods for each of the HTTP Streaming and HTTP Adaptive Streaming technologies in a mobile environment.

Table 5-1 HTTP-based streaming technologies

Technology		Specified By	Support on Android (*1)	Support on iOS
HTTP streaming		—	Supported	Supported
HTTP adaptive streaming	MPEG-DASH	ISO/IEC	Android 4.4 and higher (*2)	Not supported
	HLS	Apple	Android 4.0 and higher	Supported
	Smooth Streaming	Microsoft	Not supported	Not supported
	HTTP Dynamic Streaming	Adobe	Not supported	Not supported

(*1) The support status is based on a standard Android implementation. For the support status on smartphones, please refer to products' specifications.
(*2) Android 4.4 provides a video playback API, which will be necessary to support MPEG-DASH. The use of MPEG-DASH in video delivery requires implementation of MPD (Media Presentation Description) file analysis and processing of adaptive streaming logic using the above-mentioned video playback API on the video playback application.

5.2.1.1. HTTP adaptive streaming

In the ever-changing state of the mobile communication environment, it is suggested that HTTP Adaptive Streaming technology be adopted for its ability to provide smooth and uninterrupted video streaming and effectively utilize network resources. While HLS and MPEG-DASH are supported on Android OS and iOS, it is requested that which technology to choose be determined based on whether it is supported by the target Operating System(s) in video delivery, the content generation tools used, and the delivery server.

To take advantage of the benefits of Adaptive Streaming technology, it is suggested that a set of three video stream files in three different bitrates be prepared (in case of HLS,

four files(bitrates) because a file of an audio-only stream is also included). While the relationship between image resolution and bitrate was explained in Section 5.1., it is requested that the image resolution level in each stream be adjusted with some margins added to lower and higher rates to make a wider range of bitrate available. As a reference, the bitrates used in "d anime store" are shown in Section 6.5.

(Note) DOCOMO's 2014 summer models do not support HTTP Adaptive Streaming using video data encoded with H.265/HEVC. Therefore, it is requested that any data using HTTP Adaptive Streaming be encoded with H.264/AVC.

5.2.1.2. HTTP streaming (pseudo streaming)

For video delivery using HTTP streaming technology, it is suggested that content be made available at several different bitrates and image resolutions to allow users to choose the image quality they prefer according to the states of their network and usage. Please refer to Section 5.1 for the relationship between image resolution and bitrate. Please refer to Section 6.5 for the operation parameters used in DOCOMO's d anime store.

5.2.1.3. Download scheme

The Download scheme is highly effective to enable users to comfortably watch video in a mobile environment. Users can experience stress-free playback when they are on the road and in an unstable network environment by downloading video data in a stable network environment in advance.

5.2.2. Time zone-based delivery rate control

In Chapter 3, a drastic degradation of image quality resulting from video stall was explained. This problem occurs in a congested mobile network even when the encoding bitrate is raised to improve image quality. In order to avoid such degradation, it will be effective to perform the following delivery rate control after preparing data sets of the same video in multiple encoding bitrates. In doing so, please use the peak time zones in Section 2.1 and suggested encoding bitrates in Section 5.1 as reference.

1. Choose video data encoded in a low bitrate for a time zone where a mobile network is congested
2. Do not use an unnecessarily high coding bitrate for access from a mobile network even in a time zone where the mobile network is not congested.

5.2.3. UI (User Interface) enabling selection of delivery bitrate

To enable users to smoothly change bitrate according to their network state even in the middle of playback is one of the simple and effective means to improve quality of

experience of users. This approach mitigates users' dissatisfaction because they are given power to make a decision during video viewing as to which to prioritize, image quality or smooth playback.

The above-mentioned measure is suggested to be used in fixed –rate video delivery such as HTTP streaming and progressive download.

For an example of how to operate delivery bitrate selection and switching functions at d anime store , please refer to Section 6.2.

5.3. Other optimization measures

This section describes other suggested measures not covered in the previous sections.

5.3.1. Adjustment of still images on web sites

In video delivery service, still images on the web sites to introduce video have the second-largest volume of data following video images themselves. Although the use of rich still images is effective in introducing video, slow browsing caused by "heavy" web sites is a serious issue that can prompt users to cancel services. For the use of still images on web sites, the following are four generally-known approaches to be taken. It is strongly requested to consider their introduction in combination with your video delivery schemes to improve quality of experience perceived by users visiting web sites.

5.3.1.1. Resizing for optimal resolution

One still image is used in various sizes in web sites. When only the visual quality of image is taken into consideration, generating a reduced version of a big file will suffice. From the standpoint of disk capacity, however, maintaining multiple sizes of one still image may not be always preferred. On the other hand, sending an image in an appropriate size greatly reduces the volume of data. Data volume reduction leads not only to shorter transfer time but also to faster and smoother drawing processing on smartphones. This improves the quality of experience of users.

5.3.1.2. Image volume reduction (optimization)

One of the effective measures well known since the feature-phone age is the utilization of image volume reduction (optimization) tools. These tools can reduce the volume of images by modifying them to an extent difficult to recognize with human eyes. Data compression methods used by tools include increasing compression ratio in JPEG, use of PNG24 or PNG8 in PNG, and deletion of header information such as Exif information.

5.3.1.3. Utilization of browser cache

Utilizing browser cache, known as a measure for faster web browsing, is effective for images and large-sized JavaScript/CSS files.

Setting cache control of the browser on a web server used for delivery of images enables the same images to be displayed when users open the web page again. This is done without the need for transfer (or with a request with an if-modified-since header) or download of images.

5.3.1.4. Delayed image loading

In web pages, images are used even outside the display area. A mechanism to delay loading of those images until the time they enter the display area can shorten the loading time required to complete a screen transition.

There are some plugins publicly available (e.g., Lazy Load Plugin for jQuery), which can be used to facilitate the introduction of the above-mentioned mechanism and to improve the quality of the experience for users.

Chapter 6. Examples of measures taken for NTT DOCOMO's video site

This chapter provides specific examples of how the measures such as those described in Chapter 5 have been implemented by NTT DOCOMO to utilize its network efficiently.

6.1. Delivery technologies and their usage tendency at "d-Amine Store"

In accordance with the description in Section 5.2, "d anime store" (DOCOMO's anime store) offers two video delivery formats: one for Android and the other for iOS. HTTP streaming and Download are offered for Android and HLS and Download for iOS. In addition, the Store offers three image quality levels.

6.1.1. Relationship between delivery technologies and image quality

In general, anime users ask for high quality images. At d anime store, however, some different tendency has been observed among its users. Table 6-1 shows usage ratios for different image quality levels at d anime store. According to Table 6-1, in case of 3G/LTE, more users choose the image quality that ensures stable playback. In case of Wi-Fi, which secures sufficient throughput, they give greater importance to higher image quality. The data shown in this table indicate that users choose the delivery scheme and image quality the most suitable to their environment. It also shows that they are actively utilizing Wi-Fi to view high-quality video.

Table 6-1 Ratio of number of playback times for each usage type (d anime store, second half of 2013)

Ratio of number of playback times		Image quality			Total
		Normal	Good	Excellent	
HTTP streaming	3G/LTE	Approx. 30%	Approx. 20%	Less than 10%	Approx. 40%
	Wi-Fi	Less than 10%	Approx. 10%	Approx. 20%	Approx. 30%
Download	3G/LTE	Approx. 10%	Less than 10%	Less than 10%	Approx. 10%
	Wi-Fi	Less than 10%	Less than 10%	Approx. 10%	Approx. 10%
Total		Approx. 40%	Approx. 30%	Approx. 30%	100%

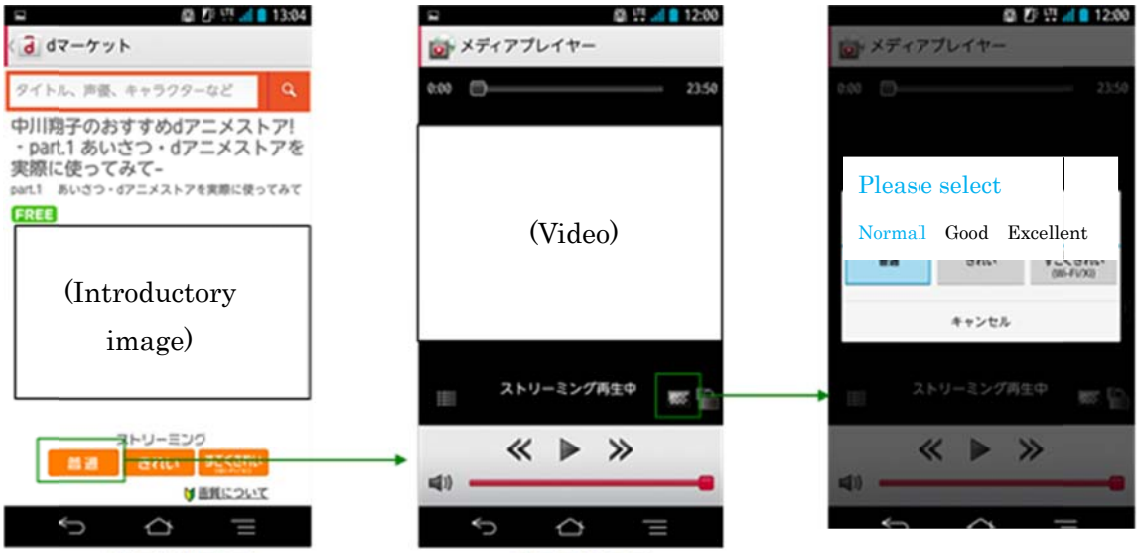
(Note) Totals may not add up due to rounding.

6.1.2. DRM license validity period for Download scheme

For Download, d anime store issues a DRM license valid for 48 hours from the time of download. This service design enables the user to play video downloaded on a previous day without generating any further communication. In addition, even after 48 hours from the download, the user can play the downloaded video only by reacquiring their DRM license file. This makes the user feel secure in using the service even when their network environment cannot offer sufficient throughput for video delivery.

6.2. UI enabling selection of delivery bitrate

At d anime store, the user is able to select the delivery bitrate on two screens as shown in Figure 6-1.



1. Video selection screen ("d-Market")

2. Video playback screen (Media Player)

3. Image quality screen (Media Player)

Figure 6-1 UI enabling selection of delivery bitrate (d anime store)

6.2.1. Video selection screen

The user can select a level of image quality from "normal," " good," or "excellent" at the time of playback. There is no default setting. The screen is designed to enable the user to select any image quality at one touch of a button.

At the same time, a link to description of each image quality is placed under all the playback buttons in order to keep the user aware of image quality.

6.2.2. Image quality selection during playback

While on the video playback screen, pressing the image quality selection button displays a pop-up screen for image quality selection. This allows the user to switch to a different delivery bitrate according to the changing network conditions during video playback. The interworking with the resume (bookmark) function enables the video to start playing from the last point the user was watching before the change of delivery bitrate.

6.3. Adjustment of still images on web sites

6.3.1. Resizing for optimal resolution

"d anime store" uses CMS(Contents Management System) to manage still pictures of video. When one image is registered, the system generates files in 3 sizes, L, M, and S. This enables delivery of data in an optimum size for each screen.

6.3.2. Volume reduction (optimization)

At its service launch, d anime store was using PNG24. After confirming that visual degradation was limited with PNG8, the store has been using a tool mainly based on PNG8 compression to perform optimization. By means of the tool, the data volume of the TOP page has been compressed to one-third of its original volume.

6.3.3. Utilization of browser cache

"d anime Store" sets the cache period of images to a relatively longer time so that the cache is still effective when the user revisits after several days.

6.4. Notices for users

As mentioned in Section 6.1, the quality of experience of users can be improved by providing video using the delivery scheme and the bitrate suitable to the specific network environment available for each user. In addition, it is necessary to encourage users to be conscious of the suitability of the delivery scheme and the bitrate to their network environment in order to promote measures to improve the quality of experience. Certain means such as a notice on a web site will be effective to make more users aware of such measures. Figure 6-2 shows a notice given for that purpose at d anime store as an example.

The mechanism explained in Section 2.2 has made regular video viewers sensitive to data volume. Therefore, they are actively utilizing Wi-Fi and selecting image quality. This is why d anime store provides information on Wi-Fi utilization in the notice shown in Figure 6-2.

Change of communication speeds for Xi

If you are subscribed to a “Packet Flat- rate Service for Xi,”, a 128-kbps cap is imposed on the your communication speed for the rest of the month when the volume of data used in the current month exceeds a certain threshold (3GB, 7GB).
If the speed is reduced to 128 kbps, you may find it difficult to view video properly. Therefore, we propose certain measures to enable you to avoid such communication speed change and enjoy your comfortable Anime Store experience.

--Use different lines at different times--
1. Utilize wireless LAN (Wi-Fi) at home.
You can download the episodes of anime you want to view all together via wireless LAN! You can view them without doing anything special within 48 hours after the download. Even after 48 hours, what you need is only to confirm the license.
2. Utilize docomo Wi-Fi outdoors
docomo Wi-Fi is available at over 70,000 locations, including train stations, airports, coffee shops, fast food restaurants, family restaurants, and convenience stores. Start now and use WiFi for free forever! You are recommended to apply now!
http://www.nttdocomo.co.jp/info/news_release/2012/08/22_00.html
3. Utilize “Normal” mode when using the mobile phone line (Xi、3G).
With the “Normal” mode, you can view approximately 180 episodes before the data volume reaches 7GB.

--Examples--
・First, you download 10 episodes or so at home or via docomo Wi-Fi.
・Within 48 hours after the download, you view them without doing anything special.
・After 48 hours, only you have to do is license authentication via the mobile phone.
・As you view more of the downloaded episodes, you download another 10 episodes.
・Only for the episodes you haven’t downloaded, you watch them in the “Normal” mode over Xi or 3G.

--Estimated data volume--
“Excellent” (data volume/episode:300MB, 3GB: 10 episodes, 7GB: 24 episodes)
“Good” (data volume/episode:100MB, 3GB: 31 episodes, 7GB: 72 episodes)
“Normal” (data volume/episode: 40MB、3GB:77 episodes, 7GB: 179 episodes)
(Note) The runtime of one episode: approx. 30 minutes. Assumptions: Normal=40MB, Good=100MB, Excellent＝300MB.

--Packet Flat-rate Service for Xi--
For details, please visit DOCOMO web page shown below.
http://www.nttdocomo.co.jp/charge/packet/

--If you want to know your current data usage--
Data volume notice service
http://www.nttdocomo.co.jp/charge/online/notification_service/about/index.html
-->You will receive a notice when your data usage in the current month has reached［2GB, 3GB］or［6GB、7GB］.
MyDocomo
http://www.nttdocomo.co.jp/charge/online/data_traffic/
-->By logging in with your docomo ID, you can see your accurate data usage status.
Data Volume Verification App
http://www.nttdocomo.co.jp/charge/online/data_app/
-->This app enables you to check an estimate of your packet data volume.

Figure 6-2 A notice introducing measures for comfortable video viewing (d anime store)

6.5. Video specifications at "d anime store"

6.5.1. File specifications for HTTP streaming/Download technologies

"d anime store" is NTT DOCOMO's video delivery service launched in July 2012.
This service is available for NTT DOCOMO's Android terminals (DRM-capable models) released in and after November, 2011. The video specifications for the Store have been determined based on the specifications supported in playback on these terminal models in accordance with the concepts shown in Table 6-2.

.

Table 6-2 Image quality concepts of d anime store

Image quality	Concept
Normal	This is the image quality that responds to the needs of users who want to watch video even if its image quality is not entirely satisfactory while on the move or in any other situations difficult to get stable throughput in a LTE/3G environment.
Good	This is the image quality that satisfies users who watch anime video in a stationary position in a LTE/3G environment.
Excellent	This is the image quality that provides users with the highest level of satisfaction through a video-viewing terminal in a stable network environment via WiFi, etc.

Based on the concepts above and the characteristics of anime content, d anime store adopts parameter specifications different from those shown in 5.1.1. Specifically, the Store uses different frame rates so that they can make the best use of limited bitrates to realize "excellent images."

Table 6-3 Video specifications for H.264/AVC (d anime Store)

H.264/AVC, AAC-LC		Image quality		
		Normal	Good	Excellent
Profile		Baseline	Baseline	Baseline
Level		3	3	3
Resolution	16:9	428×240	428×240	720×404
	4:3	320×240	320×240	640×480
Bitrate (kbps)	Video	200	500	1,500
	Audio	48	96	192
	Total	248	596	1,692
Video	Frame rate	20	20	29.97
Audio	Sampling frequency (KHz)	44.1	44.1	48
	Stereo/Monaural	Monaural	Stereo	Stereo

As the models for 2014 summer or later support H.265/HEVC, d anime Store will expand its content lineup supporting H.265/HEVC. The content specifications for H.265/HEVC have been determined in light of the concepts above to reduce the amount of data and achieve "high-speed download." The same frame rates are used for all the levels of normal, good, and excellent because verification results showed that higher frame rates had little impact on the image quality. Based on the concepts above and the characteristics of anime content, the Store adopts parameter specifications different from those shown in 5.1.2.

The following table shows video specifications of d anime store for H.265/HEVC.

Table 6-4 Video specifications for H.265/HEVC (d anime store)

H.265/HEVC, AAC-LC		Image quality		
		Normal	Good	Excellent
Profile		Main	Main	Main
Level		3	3	3.1
Tier		Main	Main	Main
Resolution	16:9	428×240	428×240	852×480*
	4:3	320×240	320×240	640×480
Bitrate (kbps)	Video	100*	250*	700*
	Audio	48	96	192
	Total	148	346	892
Video	Frame rate	29.97*	29.97*	29.97
Audio	Sampling frequency (KHz)	44.1	44.1	48
	Stereo/Monaural	Monaural	Stereo	Stereo

*Different from H.264/AVC

6.5.2. File specifications for HLS technology

"d anime store" uses HLS for video streaming for iOS in accordance with the rules of Apple Inc. The Store delivers video as a package of four files, which includes an audio-only file according to the rules of Apple Inc. in addition to the three video files shown in Table 6-3. The rules say that the size of an audio-only file "should never exceed 64 kbps even for a moment," the Store encodes the audio data with an average rate of 32 kbps so that it will not exceed the maximum rate of 64kbps at any time. This is because the file has been generated with VBR (Variable Bit Rate) due to such reasons as encoding and software-related work, although generating a 64-kbps file with CBR (Constant Bit Rate) may seem more reasonable.

The m3u8 file that integrates all the streams lists them in the order of excellent, good, normal, and audio only. This arrangement has been made to accommodate customer requests that they "want to enjoy anime with image quality as high as possible."

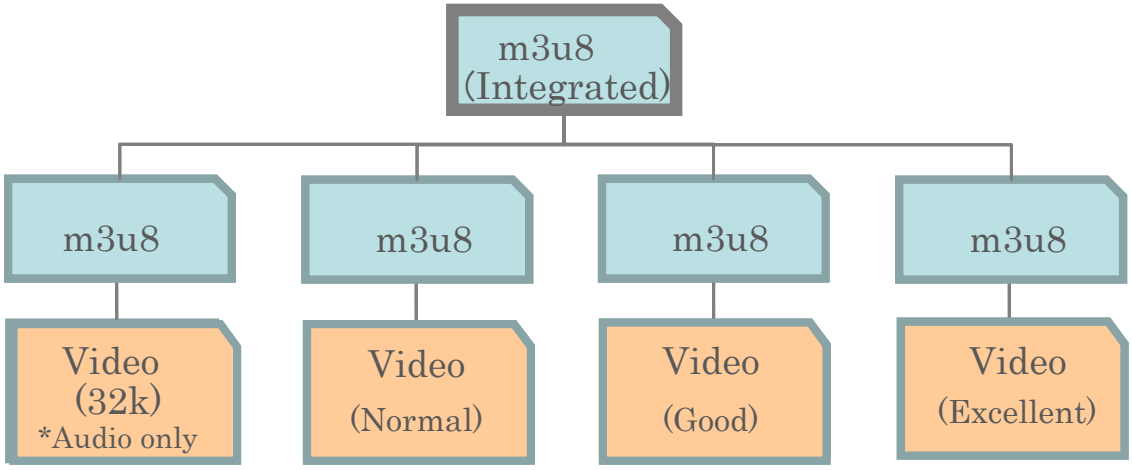


Figure 6-1 File structure for HLS (d anime store Excellent)

On the other hand, some customers were asking for a reduction in the volume of LTE communication and a capability to specify service quality by themselves. NTT DOCOMO altered the Store mechanism in April 2014 so that customers can set the highest image quality level by themselves.

- Excellent: (Excellent--> Good -- > Normal -- > Audio only)
- Good : (Good -- > Normal -- > Audio only)
- Normal : (Normal -- > Audio only)

6.6. 【Information】Introduction of tools

This section introduces some major tools that can be used to achieve the measures described in Section 5.3. Please note that the links shown below may be subject to change. Also, please be advised to use any of the tools in accordance with their terms of use or licenses specified by their respective owners. NTT DOCOMO assumes no responsibility for any damages, troubles, or other consequences resulting from the use of the tools.

Still picture volume reduction tool:
modpagespeed: https://code.google.com/p/modpagespeed/
Smush.it: http://www.smushit.com/ysmush.it/

Web page improvement tools:
PageSpeed Insights: https://developers.google.com/speed/pagespeed/insights/
Yslow: http://developer.yahoo.com/yslow/

Acknowledgements

NTT DOCOMO would like to extend our deepest appreciation to J-Stream Inc., DWANGO Co., Ltd., Forecast Communications Inc., and NTT Plala Inc. for their valuable insights provided during the creation of this document.
NTT DOCOMO is also grateful to NTT's Network Technology laboratories, Service Evolution Laboratories, and Media Intelligence Laboratories for their technical assistance provided for us in preparing this document.