Anthony Vetro
Mitsubishi Electric Research Labs

Video Delivery Challenges and Opportunities in 4G Networks

Amit Pande, Vishal Ahuja, Rajarajan Sivaraj, Eilwoo Baik, and Prasant Mohapatra
University of California, Davis

Wireless network traffic is dominated by video and requires new ways to maximize the user experience and optimize networks to prevent saturation. The exploding number of subscribers in cellular networks has exponentially increased the volume and variety of multimedia content flowing across the network. Video delivery is both an opportunity and a challenge in 4G networks such as Long-Term Evolution Advanced (LTE-A). It issues unique challenges for optimal delivery, caching, rate adaptation, quality assurance, and assessment. Such demand can quickly saturate these networks, making it difficult to promise end users acceptable quality of service (QoS) and quality of experience (QoE).

In this article, we classify mobile video applications into four broad categories and analyze the current trends, issues, and opportunities in mobile video delivery in cellular networks.

Video Traffic Characteristics

Video traffic requires special treatment relative to data traffic, not just because of its sheer volume but because of the time sensitiveness of many multimedia applications, such as live streaming and chat. The three main network QoS parameters are delay, jitter, and packet loss. Delay consists of four components: transmission, propagation, network queuing, and processing delays. Jitter indicates variations in network delay due to fluctuating network conditions. To remove jitter, some receivers introduce a de-jitter buffer (not possible in real-time applications). Different packets have different impact on quality, and late arrivals can be ignored to make space for current packets. This is typically not the case with regular data traffic.

Similarly, an equal amount of packet loss may lead to degrees of loss in the perceptual quality of two videos depending on the content, video codec, container, group of pictures (GOP), and bit rate. For example, Figure 1 shows an example of equal packet loss in different video frame types. It is evident that packet losses corresponding to I frames cause more severe degradation than P or B frames in a video.

4G technologies such as LTE-A allow increased bitrates, dedicated multicast channel for video downlink; carrier aggregation; cooperative communications; multiple-input, multiple-output (MIMO); and other enabling technologies. This makes LTE-A an attractive option to cater to future video demands. Femto cells are used to off-load traffic from base stations (evolved Node B [ENB]) in home or small business settings. Similarly, WiFi connections can also be used in hotspots such as a Starbucks or other locations for the same purpose. Figure 2 gives an overview of video delivery scenario in 4G networks, which is broadly classified into four categories:

- *Video-on-demand* (VoD). By far, VoD generated from sites such as YouTube, Netflix, Hulu, and other social networking and movie websites is the largest contributor of video traffic in wired and wireless networks. Videos posted on Facebook and some immersive and augmented video applications also fall in the same category. These videos are usually not broadcasted live (real time). To minimize the end-to-end latency

Editor's Note

Video services occupy a substantial percentage of today's network traffic. The increased bandwidth offered by 4G networks has the potential to provide some relief, but there are substantial challenges to realizing optimal delivery. This article examines performance factors for several video delivery scenarios.

1070-986X/13/$31.00 © 2013 IEEE

Published by the IEEE Computer Society

(a) (b) (c)

Figure 1. Video distortions in I, P, and B frames of a video caused by selective dropping of packets in the MAC layer (10 percent). (a) I-frame packet loss (b) P-frame packet loss, and (c) B-frame packet loss.

when delivering the videos, content servers are used to cache them. These videos are not of high quality and are transmitted using connection-oriented protocols such as TCP. Buffering and other mechanisms are used to improve the user experience.

- *Video multicast*. Video multicast is gaining attention from industry to provision real-time streaming of events such as a soccer match or transmission of HDTV over 4G networks using a dedicated Multimedia Broadcast Multicast Service (MBMS). It typically uses the User Datagram Protocol (UDP) and is sensitive to network packet losses, which cause visual impairments in the form of frame blocking, blurring, and freezing.
- *Video chat*. Interactive video chat between two or more parties is gaining momentum. Such chats occur in real time, use UDP, and are extremely sensitive to the end-to-end delay. Typically, low resolutions and frame rates are supported.
- *Video uplink*. People upload videos captured using smartphone cameras to social networking site such as YouTube. Video uplink usually does not require a live, real-time feed, and reliable connection-oriented protocols can be used.

In the following sections, we give an overview of applications that fall into these four categories and discuss the major issues.

Video-on-Demand

Video-on-demand forms a large portion of the current video traffic, with the current players being video sites such as YouTube, Facebook, Hulu, and Netflix. According to a 2010 study,[1] 35 hours of videos were uploaded to YouTube every minute, and there were more than 700 billion playbacks. Netflix has more than 23 million subscribers in the US and Canada and accounts for 29.7 percent of the downstream traffic in the US.

Considering these statistics, we can see that the design and traffic management decisions taken by these video providers will have a significant impact on the overall networking infrastructure. HTTP and TCP are the default protocols used by most of the video streaming services to stream data to clients via single or multiple content delivery networks (CDNs), such as Akamai and Limelight. Also, HTTP streaming is well established, which means that the CDNs can ensure that the service can reach clients through network address translators (NATs), and they can do so in a cost-effective manner.

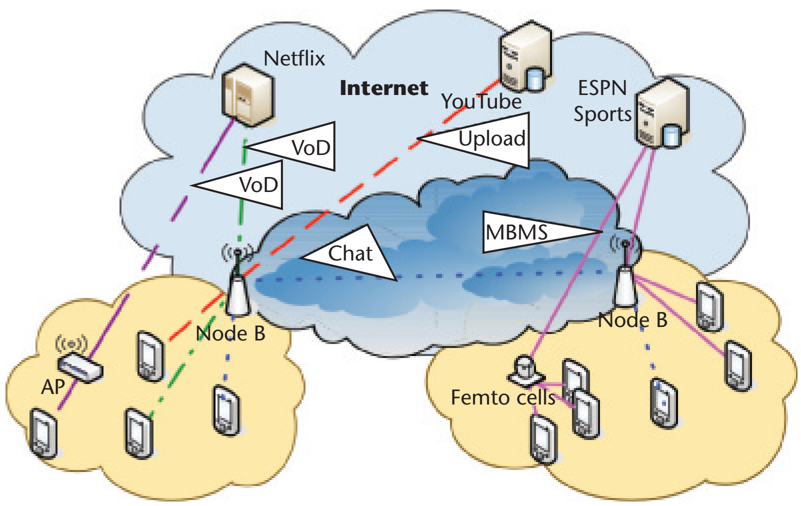

Figure 2. Block diagram of a video delivery scenario in 4G networks. Video-on-demand (VoD) sites such as YouTube and Netflix deliver content via various channels and services, including Multimedia Broadcast Multicast Services (MBMS) and Femto cells.

Characterization

Mobile devices such as smartphones do not buffer an entire video, so the video is progressively downloaded in multiple chunks. The most popular delivery mechanism is *progressive download*, which account for 60 percent of the cellular traffic.[2] During progressive download, a single video is downloaded with a single HTTP request for the entire object from the client. Sometimes each chunk has a separate HTTP connection with a specific byte-range request. For example, mobile YouTube uses progressive download with multiple byte range requests. Cellular traces can have a lower file size and video quality than WiFi traces on same device.

The increased popularity of progressive download strategy is because of its simple implementation; no dedicated streaming server is required and the video is transmitted through a standard HTTP port. Efficient and adaptive strategies can be used for video delivery, such as the HTTP Live Streaming (HLS) or MPEG Dynamic Adaptive Streaming over HTTP (DASH) protocols.[3] They can be used to deliver the same content to different screens and adaptively adjust bit rates according to the end users.

Rate Adaptation

Rate adaptation in mobile videos is an open issue. T. Brandon and H. Johari discovered the confusing nature of rate adaptation used by existing video streaming algorithms in Hulu, Netflix, and Vudu.[4] In almost all cases, the bandwidth used by the service does not closely follow the available network bandwidth, leading to a reduced throughput.

Rate adaptation can be done using scalable compression techniques. Quantization, frame rate, and screen resolution are the three dimensions that can be adapted and scaled using existing codecs. However, scalable codecs (such as H.264 SVC) incorporate significant overhead in compression performance that is unacceptable in wireless and cellular scenarios. Thus, a preferred solution is for the server to store multiple copies of the video at different qualities and transmit the appropriate stream to different clients according to their network condition. Any rate-adaptation algorithm must be able to detect the available network traffic (last hop) and check the availability of computational resources at mobile device.

User Behavior

Only a fraction of the downloaded or streamed videos are actually watched by end users. This is largely due to the tendency of users to skim through a video or click the recommended links even before watching the entire video.

Thus, progressive download techniques lead to significant overhead in such cases because the service typically keeps downloading the remaining portions of video (unless rate limited by the memory capacity of a smartphone). Adaptive streaming techniques such as DASH don't buffer the video at the client, thus avoiding wasted network resources due to unviewed downloads.

Caching

Caching is a popular technique for temporary storage to reduce bandwidth usage, server load, and perceived lag. Contrary to the microprocessor caches that store the most frequent content in the smallest level 1 caches, the level 1 caches used by CDN operators (closest to users) are the largest in size. This makes sense considering the vast variety of video traffic.

An interesting option would be to use scalable caching mechanisms. The base layer quality of all videos can be stored in level 1 caches, and enhancement layers for less-frequent caches can be stored in level 2 or level 3 caches. On user request, a connection may be established to the level 1 cache showing a low-quality video, which can then be transferred to level 2 or 3 caches seamlessly.[5] This hierarchical management may be helpful in reducing the caching overhead while guaranteeing a low response time for user requests.

Video Multicast

Traditionally, cellular networks have used unicast mode for video communications. However, video broadcast is gaining attention from cellular providers for live event streaming, leading to tremendous savings in network capacity.[6]

Content Scheduling

Designing an optimal link adaptation and scheduling scheme for multicast is an open problem. Mobile TV broadcasting has been recently studied with the goals of maximizing the video quality using offline measurement metrics such as peak signal-to-noise ratio (PSNR) and energy efficiency in receiver devices by considering burst transmissions and power-off modes. Using burst transmissions, service providers such as AT&T and Verizon can serve numerous clients rather than allocate dedicated resources to a each multicast group. Different constraints have been considered, such as

- maximizing the network goodput (application-level throughput),
- considering the variable channel conditions of end users,
- considering the limited resources of mobile receivers, and
- accounting for the heterogeneity of mobile devices.

However, the problem is proven to be NP-hard and heuristic solutions have been proposed.

Earlier research formulated the problem with the following service objectives:[6]

- Ensure mandatory base layer video quality to all possible users.
- Opportunistically provide higher enhancement layers for better quality.

Table 1 shows that spectrum aware assignment can achieve a significant improvement over general opportunistic assignment can be observed as the number of users in a cell increases. Uniform traffic was considered for NS-3 simulations with up to 100 types of user equipment (UE) and up to five UEs per group. The proposed scheme outperforms the opportunistic scheme by 12 to 25 percent in base layer fractions and more than 50 percent in the enhancement layer fraction.[7]

Table 1. Achievable cell throughput for user equipment groups in LTE Evolved Multimedia Broadcast Multicast Services (eMBMS) traffic using spectrum aware assignment.[7]

Number	Base layer		Enhancement layer	
of users	Proposed	Opportunistic	Proposed	Opportunistic
30	19.85	17.885	51.12	47.76
60	14.22	11.78	27.35	19.47
90	10.74	8.82	14.20	8.01

Resiliency

Multicasting implies an absence of feedback mechanisms. (Channel quality index [CQI] can be effectively used only in unicast.) The group must operate in worst-case channel conditions—that is, choose the modulation and code rate suitable to worst channel conditions experienced by a user of multicast group.

Raptor codes are used in the application layer to add redundancy to source data (video). Automatic repeat request (ARQ) requires the receiver to request the retransmission of lost or corrupted packets by means of negative acknowledgement, positive acknowledgements, or timeouts. This scheme has signaling overhead in poor network conditions and leads to full reconstruction in the event of packet loss.

Forward error correction (FEC) schemes, on the other hand, add some error-correction code to data allowing reconstruction in case of packet losses. Hybrid-ARQ, which is used in the MAC layer of LTE-A, reduces the transmission overhead of ARQ by retransmitting only FEC data instead of the entire packet. This makes HARQ perform as well as ARQ in good conditions and provide good resiliency in poor network conditions.

The coexistence of Raptor codes and HARQ leads to inefficiency in transmission, but this problem has been addressed in recent work,[8] where a joint-probabilistic model can be used to choose correct Raptor code rates and HARQ levels, reducing the transmission overhead by 10 to 15 percent for single and multiuser situations.

Cross-Layer Design

Mobile video's inability to handle wireless interference and errors was addressed in a recent work by redesigning the protocol stack to act as a linear transform—that is, the transmitted video signal is linearly related to pixels' luminance.[9] Noise perturbations are thus interpreted as a coarsening of transmitted signal samples. The authors report promising improvements: a 5.5 db gain over MPEG4 transmitted mulitcast video and resilience to user mobility. Similarly, another work introduced a mechanism for smooth video transmissions in the presence of packet losses.[10]

These works can potentially be translated to cellular networks, but all of them require the protocol stack to be customized.

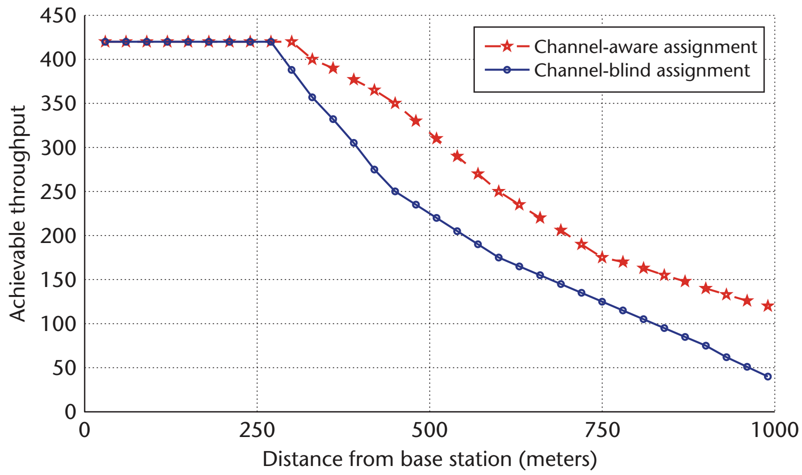

Figure 3. Improvement achieved by channel-aware schemes over channel-blind scheme. Channel-aware schemes obtain approximately a 57 percent improvement in throughput for cell-edge users with an overall improvement of 41 percent.[11]

Video Upload

Uplink carrier aggregation can significantly improve the throughput rates achievable in LTE-A networks when end users upload comprehensive multimedia content that consists of images, music, and videos. This service does not stress the network because there are no real-time requirements.

Prioritization of Edge Users

Cell-edge users suffer the most from exhaustion of resources, higher fading losses, and lower signal to interference plus noise ratio (SINR) values (and hence higher power consumption in uplink). Channel-agnostic radio resource management (RRM) further impedes their performance, leaving behind the least-satisfied traffic requirements.

Unlike the downlink scenario where the base station can increase the transmit power, smartphones with battery limitations can't increase their transmit powers to overcome the higher fading losses. Grouping of users and subsequent edge-prioritized RRM can lead to significant savings in the user battery requirements and improve the overall network throughput.[11]

Figure 3 shows the improvement obtained using a channel-aware assignment, particularly for cell-edge users who are farther from the base station. Further improvements can be realized by grouping and prioritizing users according to their channel characteristics.

Power Savings

Although video uplink should be completed within a reasonable time, a user might appreciate if the transfers also placed a minimal cost on battery life.

Generally, cell-edge users have high power consumption to increase the transmission power to compensate the low SINR values and lossy channels. Our work on edge-prioritized channel resource allocation empowers edge UE to maintain a significantly low transmission power by choosing low channels with low path losses.[11]

Bartendr is an approach to defer communication, where possible, until the device moves into a location with better signal strength.[12] The approach relies on efficient scheduling of data communications to save battery power and is applicable both to VoD and uploads. A savings of up to 10 to 60 percent was obtained for VoD deadlines. Lighter deadline constraints in video upload will particularly yield significant gains in battery efficiency.

Video Chat

Mobile video telephony or chat has gained traction in recent years with a number of offerings in the market from players such as Skype, Google Hangout, Fring, and Apple's FaceTime and iChat.

Video telephony has stringent requirements on a network with an acceptable end-to-end delay of around 150 ms (including time for encoding and decoding, transmission),[13] and it can easily saturate 200 to 1,000 kbps of bandwidth. The requirements for real-time communications require strict minimum bandwidth guarantees. As a result, UDP is used instead of TCP. This also necessitates the use of efficient low-complexity video codecs.

Architecture

Current service providers use different architectures for video telephony. For example, iChat uses a peer-peer architecture, Skype uses a hybrid architecture, and Google Hangout uses a client-server architecture.[14] The choice of architecture is crucial for good QoS in cellular networks. An architecture leaning toward the client-server model will be beneficial in cellular networks because it empowers network providers to fine tune the parameters and is consistent with the design of cellular networks.

The choice of appropriate video coding is also crucial for video chats. Multiparty chats may require different quality levels across different users having different channel quality. However, scalable encoding has significant computational overhead, which leads to a quick draining of the battery levels on a smartphone device. Transcoding at the server is another possible choice.

Resiliency

The two basic error-correction mechanisms are ARQ and FEC. A more intelligent HARQ, based on selective dropping of less important video data, can be employed to improve the performance in the video chat scenario. Although scalable codecs are default choices for provisioning multicast groups, the choice of the H.264 SVC codec, which leads to a bandwidth inefficiency of up to 30 percent, imposes a significant penalty for mobile carriers.

The upcoming HEVC codec aims to substantially improve coding efficiency compared to AVC High Profile. The goal is to reduce bitrate requirements by half with comparable image quality, at the expense of increased computational complexity, and a scalable extension is under development. Depending on the application requirements, HEVC should be able to trade off computational complexity, compression rate, robustness to errors, and processing delay time.

Discussion

We have detailed some challenges in delivery of multimedia content over 4G networks for several application scenarios. Emerging immersive applications, augmented reality, and others can also be classified into these same categories. The choice and performance of video service is largely affected by the type of application and network conditions. Resiliency to packet loss and power-efficient transmissions are important for all traffic classes. Another important challenge is the design and implementation of robust video quality assessment metrics for mobile scenarios. To augment the increasing demand for video applications in cellular and wireless traffic, these challenges must be efficiently addressed. **MM**

The choice and performance of video service is largely affected by the type of application and network conditions.

References

1. A. Finamore et al., ''YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience,'' *Proc. ACM SIGCOMM Conf. Internet Measurement,* ACM, 2011, pp. 345–360.
2. J. Erman et al., ''Over the Top Video: The Gorilla in Cellular Networks,'' *Proc. ACM SIGCOMM Conf. Internet Measurement,* ACM, 2011, pp. 127–136.
3. I. Sodagar, ''The MPEG-DASH Standard for Multimedia Streaming over the Internet,'' *IEEE Multimedia,* vol. 18, no. 4, 2011, pp. 62–67.
4. T. Brandon and H. Johari, ''Confused, Timid, and Unstable: Picking a Video Streaming Rate Is Hard,'' *Proc. ACM SIGCOMM Conf. Internet Measurement,* ACM, 2012.
5. V.K. Adhikari et al., ''Reverse Engineering the YouTube Video Delivery Cloud,'' *Proc. ACM SIGMETRICS Joint Int'l Conf. Measurement and Modeling of Computer Systems,* ACM, 2011, pp. 137–138
6. V K. Fitchard, ''Can LTE Broadcast Dam the Mobile Video Deluge?'' blog, 10 Jan. 2013; http://gigaom.com/2013/01/10/can-lte-broadcast-dam-the-mobile-video-deluge.
7. R. Sivaraj, A. Pande, and P. Mohapatra, ''Spectrum-Aware Radio Resource Management for Scalable Video Multicast in LTE-Advanced Systems,'' *Proc. IFIP Networking,* 2013.
8. E. Baik, A. Pande, and P. Mohapatra, ''Cross-Layer Coordination for Efficient Contents Delivery in LTE eMBMS Traffic,'' *Proc. 9th IEEE Int'l Conf. Mobile Ad Hoc and Sensor Systems,* IEEE CS, 2012, pp. 398–406.
9. S. Jakubczak and D. Katabi, ''Softcast: One-Size-Fits-All Wireless Video,'' *ACM SIGCOMM Computer Comm. Rev.,* vol. 40, no. 4, 2010, pp. 449–450.
10. J. Wang and D. Katabi, ''Chitchat: Making Video Chat Robust to Packet Loss,'' tech. report, MIT-CSAIL-TR-2010-031, Aug. 2010.
11. R. Sivaraj et al., ''Edge-Prioritized Channel- and Traffic-Aware Uplink Carrier Aggregation in LTE-Advanced Systems,'' *Proc. 13th Int'l Symp. World of Wireless, Mobile and Multimedia Networks* (WOWMOM), IEEE CS, 2012, pp. 1–9.

12. A. Schulman et al., ''Bartendr: A Practical Approach to Energy-Aware Cellular Data Scheduling,'' *Proc. 16th Ann. Int'l Conf. Mobile Computing and Networking* (MobiCom), ACM, 2010, pp. 85–96.
13. J. Jansen et al., ''Enabling Composition-Based Video-Conferencing for the Home,'' *IEEE Trans. Multimedia,* vol. 13, no. 5, 2011, pp. 869–881.
14. Y. Xu et al., ''Video Telephony for End Consumers: Measurement Study of Google+, iChat, and Skype,'' *Proc. ACM Conf. Internet Measurement,* ACM, 2012, pp. 371–384.

Amit Pande is a research scientist in the Department of Computer Science at the University of California, Davis. Pande has a PhD in computer engineering from Iowa State University. Contact him at pande@ucdavis.edu.

Vishal Ahuja is with the Department of Computer Science at the University of California, Davis. Contact him at vahuja@ucdavis.edu.

Rajarajan Sivaraj is with the Department of Computer Science at the University of California, Davis. Contact him at rsivaraj@ucdavis.edu.

Eilwoo Baik is with the Department of Computer Science at the University of California, Davis. Contact him at ebaik@ucdavis.edu.

Prasant Mohapatra is a professor in the Department of Computer Science at the University of California, Davis. His research interests include wireless networks, sensor networks, Internet protocols, and QoS. Mohapatra has a PhD from Pennsylvania State University. Contact him at pmohapatra@ucdavis.edu.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*