Mobile Video Ecosystem & Geo Fencing for Licensed Content Delivery

November 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The growth of online video continues to be nothing short of explosive—and not just in terms of the amount that people are watching. It's also growing in terms of throughput requirements, thanks to 3D and increased resolutions such as ultra-high definition (UHD). These two growth trends mean video's demand on mobile networks will increase exponentially and indefinitely. In fact, it would be possible to predict that in the future, the majority of all resources on the Internet, from bandwidth to controls, could be consumed primarily by just one form of media: video. Cisco's Visual Networking Index forecast that mobile video will be 78 percent of all consumer internet traffic by 2021, up from 60 percent in 2016.[1]

This paper, *Mobile Video Ecosystem and Geofencing for Licensed Content Delivery*, focuses on two critical issues that play significant factors in online digital video delivery: network and codec optimization and geographical filtering (geofencing) for licensed video content. The first part explores network and codec optimization. New delivery technologies improve the user experience while managing the required bandwidth to deliver a quality video experience. Meanwhile, new video codecs are at least 20 percent to 30 percent more efficient than legacy codecs.

The paper's second part focuses on the use of geographical filtering (geofencing) for delivering licensed video content subject to geographical constraints under licensing agreements. It explores aspects such as the risk of location spoofing and the need for a trusted source to verify each device's location. One challenge is the current lack of standards for communicating trusted network-based device location for a roaming device.

In today's fast-paced world of telecommunications innovations, service delivery and data consumption, video services have been front and center of bandwidth utilization. Video traffic has been growing at a rapid pace and there is no sign of it slowing down.

Even with the advancement in codec developments where lower bit rates are expected to achieve the same resolution, video traffic is still gaining momentum. In a white paper released by Cisco about its' Visual Networking Index in June of 2016, it is forecasted that mobile video traffic will be 78 percent of all consumer internet traffic by 2021, up from 60 percent in 2016.[2]

On the other hand, there are finite and scarce radio network resources that be must utilized in the most efficient way. In the wireless telecommunications world, there is always a need to strike a balance between saving network resources and quality of experience. Mobile video service is not exempt from this rule. In fact, due to the amount of traffic video brings into the network, it is where optimization is most needed. .

The white paper highlights what is happening around the industry from the perspectives of both the video content service provider and the mobile operator. For video content services, there are new delivery technologies being developed to improve the user experience while managing the required bandwidth to deliver a quality video experience. There is also a more rapid adoption of new video codecs, which are at least 20 to 30 percent more efficient than legacy codecs.

Mobile operators will also have more flexibility going forward with new standards that will allow policy decisions to be enforced in near-real time – either by the mobile operator or by the video service provider. The application service providers and mobile operators will need to work together to get the most out of the new standards and technologies, and also provide the best experience while protecting the network under

[1] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
[2] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

times of congestion. These topics are covered in the following sections by briefly reviewing the current technical landscape and providing some recommendations.

Another topic addressed in this paper is delivery of licensed video using geo-fencing. The concept of geo-fencing has been around for decades. Typically, geofencing applications target small areas in the vicinity of a Point-of-Interest (POI). The basic idea is to draw a virtual perimeter around a POI (or POIs) and initiate a preset action, based on alerts sent by a geo-fence application, when a mobile device either enters or leaves the geo-fenced area. While commercial uses of geo-fencing by retailers are easy to conceptualize, and have been in use for a long time, many other uses, such as fleet management by a trucking company, demonstrate the breadth of market reach of this technology. Nonetheless, the use of geofencing for licensed video content delivery has, to the best of our knowledge, not been previously addressed.

In the U.S., as in many other countries, most video content, whether broadcast or streamed online, is subject to strict regulations (license) that determine where such content can be received and consumed. We are all familiar with FCC license necessary for a broadcast TV station to operate. While issues related to licensing compliance for broadcast TV, intended primarily for fixed home-based users, have been successfully addressed for decades, today's mobile users consuming streaming video online present a significant new challenge. A mobile user, by definition, is not confined within the licensee's geographical distribution boundary. A mobile could be consuming content in Region B when the content is only licensed for Region A. Addressing this problem is the focus of the second part of this paper.

In the context of this paper the common understanding of the term 'geo-fence' has been expanded from its usual connotations of buildings or city blocks to include very large geographical regions. The regions of interest could range anywhere from a collection of postal codes to city/cities, state/states, or in some cases even the entire country (e.g., in Europe). However, in spite of the expansive view of the geographical areas involved the underlying principles of geofencing remain unchanged – a perimeter encircling the geographical area of interest is defined and action is taken (or not taken) based on a mobile's presence within or outside the perimeter. The geofencing discussion begins with brief backgrounds on geo-fencing as a concept, Content Delivery Network (CDN) architecture used by online video content service providers and key aspects of licensed content delivery in the U.S. This is followed by an overview of LTE Location Architecture and exploration of how the location of a mobile device – key input for any geofencing function – may be determined. The risk of location spoofing and the need for a trusted source of device location is emphasized. The paper notes the current lack of standards for communicating trusted network-based device location for a roaming device and concludes with recommendations on how one may proceed towards a solution. The discussion emphasizes the need for close collaboration between all stakeholders, especially the mobile operators, content owners, content distributors and 3rd party aggregators. It is hoped that the recommendations would be used to drive activity in Standards Development Organizations such as Third Generation Partnership Project (3GPP) or Open Mobile Alliance (OMA) for developing standards-based solutions with global industry support. Only with broad global adoption of a standardized solution, vastly simplifying future implementations and upgrades that would surely be necessary, can a realistic solution to the problem can be achieved.

1. INTRODUCTION

Mobile video is the fastest growing traffic on LTE networks and it is projected to account for more than 6 exabytes of data delivered per day by 2020, a 46 percent year-over-year increase. Mobile video streaming will represent 79 percent of all mobile data traffic by 2020.[3]

There are many factors contributing to this strong growth:

- Large-screen smartphones with better battery life, display characteristics and processing capabilities are replacing first-generation devices
- LTE is being used as a substitute or complement for fixed broadband
- There is a growing availability of high-quality video content for mobile consumption, as well as new services, such as virtual reality (VR) and 360-degree video, which stimulate usage and increase traffic
- Data-inclusive services by mobile operators create even more demand for video
- User-generated video content popularity
- The evolution to 5G networks is expected to support data rates to tens of megabits per second, again increasing video consumption

Despite the still relatively small form factor of mobile devices, mobile video consumption is becoming an important part of the changing television landscape, and viewers are using their smartphones to augment the primary screen. In a recent study,[4] 25 percent of respondents watched more than 2 hours of video on their smartphone every weekend.

While consumers love the convenience of mobile video, the experience can sometimes be marred by slow start times, buffering and high data usage. Not surprisingly, millennials are the mobile video "power users." In the same study, 30.6 percent of millennials said they watched more than 2 hours of video per week, while 11.7 percent watched more than 5 hours per week and were more likely to use their data connection while doing so. Operators are beginning to adopt a "mobile-first" strategy to reach millennials who are twice as likely to watch video on their mobile devices as they are on a TV.

Despite the growing amount of premium content available for mobile, short video clips and user-generated content remain popular. However, consumption of long-form content, particularly scripted and episodic content, is rapidly gaining traction. While YouTube is still the predominant source of on-demand content for mobile viewing, premium over-the-top (OTT) has seen incredible growth over the past few years, driven mostly by the globalization of services such as Netflix (Figure 1).

[3] *Who will satisfy the desire to consume?"* Nokia Bell Labs Mobility Traffic Report. 2016.
[4] *Mobile Video: Exposed*, SVA. December 2016.

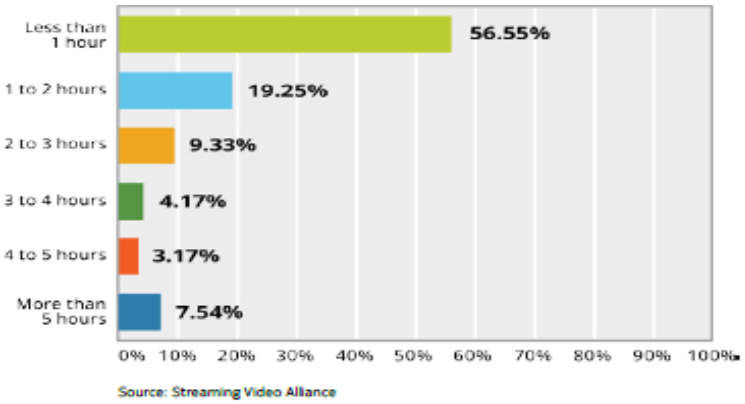

Figure 1: Sources of Most Watched Video on Smartphones.

Another point to mention is that "live is not dead." Facebook and Twitter are driving a live online video explosion, outbidding traditional channels for the rights to stream content such as major league sports, demonstrating that social media and live video are a potent combination (Figure 2). Live gaming is also a popular genre, with Twitch delivering over 2 million streams every day.

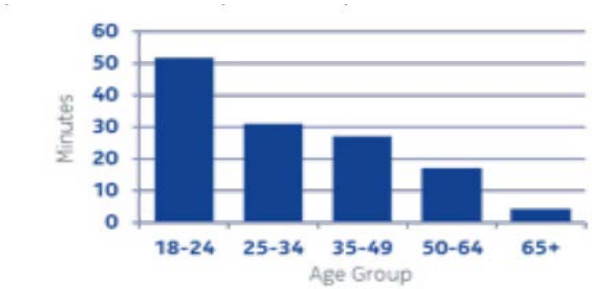

Figure 2: Weekly Consumption of Live Mobile Video in Minutes (2016).[5]

Virtual Reality (VR) and 360-degree video are also expected to be important drivers of mobile video consumption. So-called light mobile (e.g., Google Cardboard) and premium mobile (e.g., Gear VR) headsets already make use of a consumer's mobile device as the viewing mechanism. New encoding and delivery techniques such as Viewport Adaptive Streaming have the potential to deliver a high-quality VR experience across a mobile network, provided the network can manage fast transport of updates when a user changes viewing direction.

TV providers are looking to expand the footprint of their offerings. Today they are mostly constrained to "big-screen" experiences within the reach of fixed-line broadband access networks, complemented by "second-screen" mobile services. The introduction of Multi-access Edge Computing (MEC) and the evolution to 5G will change this landscape. Faster, lower-latency mobile networks with embedded IP video awareness expand the potential scope of service offerings, enabling TV providers to offer big-screen TV experiences to new subscribers and in previously unreachable locations. MEC and 5G enable mobile networks to deliver content at a quality that has previously been unachievable, and to devices where fixed-line video delivery is traditionally assumed to be the default.

Large service providers often have fixed and wireless service offerings where the wireless footprint can be up to 300 percent greater than the fixed-line services footprint, although the service footprints inevitably

[5] *Nielsen Total Audience Report*. Q2 2016.

overlap. By enhancing mobile networks with embedded IP video awareness, service providers can either expand existing service coverage or create entirely new products and services, including premium TV services delivered to big screens using the mobile network.

In this paper, the term “geofence” has been expanded from its usual connotations -- such as a building, city block, sports arena, shopping mall -- to include very large geographical regions. The key input necessary for a geofence to function is the mobile’s geographical location.

We are familiar with the ease and frequency with which today’s mobiles locate themselves. All devices come equipped with, not just GPS, but other positioning technologies such as sensors or image recognition. For many smartphones, the location function runs continuously in the background. Thus, it would be easy to conclude that the central task of any geofencing implementation (i.e., locating the mobile device) would be a straightforward task. However, for licensed video, this is not the case for one overarching reason: trust. The data originating in the mobile device, irrespective of the device OS, cannot be trusted. While this fact is of no consequence for normal mapping-type applications where occasional errors/glitches are always tolerated, the situation involving a legally binding license is entirely different. A source of device location that can be trusted is needed by the geofencing function. This reliable source of mobile device location is the mobile network that serves the device.

The paper delves into various options, with their pros and cons, for determining network-based mobile device location. The primary focus are the cases where video is streamed over LTE and/or Wi-Fi. Also addressed are the specific, and important, scenarios when the mobile device roams from its home network into an administratively distinct visited network. The roaming scenario is important because frequently the large geographical areas covered by licensing agreements span over multiple independent networks. In such roaming cases, while continuity of service remains the desired service characteristic, adherence to licensing agreement must remain the legally binding requirement.

2. VIDEO DELIVERY CHALLENGES AND OPTIMIZATIONS

5G’s low latency, high reliability and high throughput will spark innovation and new use cases for mobile video. With this transition, we expect growing pains such as the transition from 3G to LTE and from progressive download to ABR streaming.

Let’s take a step back to when LTE was first introduced in the U.S. market. Prior to LTE, video over mobile was delivered primarily by low-resolution progressive downloads. The progressive download behavior was well known by both the carriers and the services, and was often optimized by both the service and the carrier through some sort of pacing. This was also before the migration from HTTP to HTTPS so optimization platforms could determine the necessary pacing needed to ensure a smooth playback while managing wastage (video downloaded, but not consumed).

With its promise of higher throughput, lower latency and better efficiency, LTE saw an explosion of video consumption over mobile networks. With this, some top video services moved from a low default bitrate to a very high default bitrate as they transitioned to ABR. The belief was that the ABR client would manage the flow and adjust the quality with available capacity.

Several suboptimal policies and behaviors emerged during this transition that created unnecessary traffic on the mobile network. Work is still ongoing to address these inefficiencies (examples below), and carriers are beginning to implement their own forms of managing video traffic.

- Very high default bitrates

- Inefficient and static read-ahead buffers
- Random system-level bugs

With the transition from 4G to 5G and the explosion of new technologies and services, there could be growing pains like what was seen with the move from 3G to 4G. Because video consumes so much volume, small missteps can sometimes have large impacts.

While mobile video is often discussed as a homogenous entity, it is very diverse in application and practice. These differences come in many forms: type, delivery technology, available and default quality, codec, read-ahead buffers and encoding recipes, to name just a few. Traditional optimization techniques such as video pacing and shaping are no longer a one-size-fits-all solution. The move to encryption and to new protocols also makes it nearly impossible for mobile operators to apply these traditional approaches fairly across services. This can lead to a degraded user experience, inefficient use of network resources and/or unfair policies across services.

New delivery technologies, codecs and optimizations are being developed and deployed to address the changing mobile video landscape outside of traditional methods. These will enable new services, provide for a more seamless user experience, and create methods to enable a more inclusive mobile video ecosystem where mobile operators and application services can work together to ensure more customers can be served by scarce radio resources.

The following section is divided into two parts. The first reviews what mobile service providers and the overall wireless industry are doing to help enable new technologies while managing total tonnage and user experience. The second part focuses on standards and options that are available to mobile operators.

It is the objective of the paper that both application service providers and mobile operators are more familiar with what each side is doing to ensure that a more inclusive ecosystem can be developed. This will allow for greater efficiencies in video delivery, better management of growth and limiting congestion scenarios, all of which benefits mobile operators, application services and customers

2.1 DELIVERY TECHNOLOGY

The two primary methods of video delivery today are progressive download and Adaptive Bitrate streaming (ABR). ABR dominates mobile streaming today, but there are still valid use cases for progressive download that will keep it relevant. While these two technologies are relatively mature, there are delivery advancements that will enable new services and simultaneously place different demands on the mobile network.

Progressive download video consists of a single resolution being downloaded as a single file. The resolution selected at the beginning of play, either by the user or service, is the resolution downloaded and played for the entirety of the viewing session. In other words, the quality is static once playback starts and does not adjust down or up with varying network conditions. Progressive downloads are used for offline and pre-fetching content, and by some social media services.

As its name implies, ABR adapts the resolution based on current network conditions. To enable this, ABR encodes the video at different rates, many of which would typically be at different resolutions. Each video corresponding to a rate profile is further broken down into chunks, often 2 to 10 seconds each. This allows profile selection to be adapted at every chunk boundary, as needed. Typically, the client monitors the available bandwidth and its playout buffer levels, and makes decisions at every chunk boundary regarding what rate profile to request. When network congestion increases, a client would typically switch to a lower

rate profile to avoid a freeze frame caused by a buffer stall. Similarly, when network congestion eases (detected at the client by seeing increased available bandwidth), the client can switch to a higher rate profile.

Even with ABR, there is an upper limit to the bitrate and quality of video available for streaming to mobile devices. This upper limit varies by Over The Top (OTT) service and device capabilities, but is generally sufficient to provide a clear picture for most content today. However, this upper limit may not be sufficient for some current and future video types. This upper limit is evident for video types such as 360-degree video, which can have high clarity in its uncompressed format, but can be washed out when re-encoded and streamed by an OTT service. As Figure 3 illustrates, the left image is 1080P standard video and the right image is a 1440P 360-degree video. The standard video is clear and has bright colors whereas the 360-degree video is dark and dull. The contrast between the two is quite striking even though the 360-degree video has a higher resolution and is more than double the bitrate of the non-360-degree video.

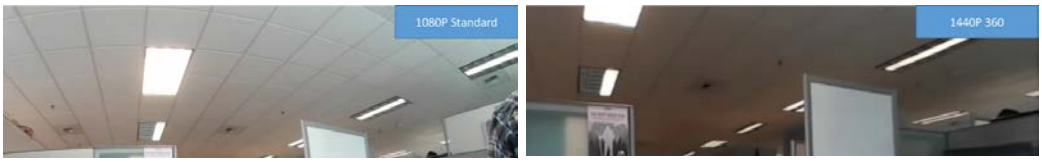

Figure 3: 1080P Standard Compared to 1440P 360 Degree.

A few novel approaches have been devised to address the challenges demonstrated in Figure 3, which are illustrated in Figure 4. The focus is on the 360-degree video; however, this is to demonstrate the current innovation.

In Figure 4, the left image is a standard 360-degree video image, which is an equirectangular projection. Like a flat map of the world, this includes redundant information around the edges as it is stretched to fit the shape. For example, maps in grade school would make Antarctica and Greenland stretch to be much larger. The same concept is employed. Because this includes redundant information, it requires the redundant bits to be encoded and then delivered, which increases the required bitrate or decreases the visual quality.

Facebook developed a solution to eliminate the redundant information by creating a cube map of the video.[6] As the name implies, this method splits the video into cubes, which can reduce up to 25 percent of redundant information from the original 360-degree video. Figure 4 shows an example of a cube map. The images are not stretched and do not contain redundant information. It also visually demonstrates how much of the image can be saved with this technique.

Figure 4: Equirectangular 360-degree Projection vs. Cube Map.

[6] https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/.

Another development is Viewport Adaptive or Field of View Adaptive streaming. This divides the video into different viewing ports, and only delivers the viewport where the user is currently looking. If the user quickly changes direction, the client/server must quickly adjust to a new viewport.

While ABR and new encoding and delivery techniques will enable innovation, they will also create different demands on the network. Viewport Adaptive streams will require very low latencies, which is much different than the video requirements of today. This is where 5G's low latency will help foster new innovative services.

2.2 VIDEO CODECS

Another form of service optimization is development and adoption of video codecs, which are used to compress and decompress video. H.264 is the most widely used codec on mobile networks today and is supported in all modern mobile devices. New codecs have been gaining momentum over the last few years thanks to growing hardware and service support. These new codecs, H.265 and VP9, promise both quality improvements and data savings.

VP9 has seen wider adoption due to the two largest mobile video players supporting the codec for playback on devices with hardware decoders. Even with support growing for the new codecs, there has been slow adoption with the exception of YouTube until early 2017.

Currently there are at least three major services streaming content to mobile devices with one of the new codecs. Google has all of its YouTube content available in VP9. Netflix has a larger percentage of its content available in VP9 and its HDR content available in H.265. Amazon has some content available in H.265. These services have the content available in both the new codecs and legacy H.264 for backwards capability. The decision about which codec to stream is a combination of the availability of hardware decoder support and a whitelisting mechanism on the OTT side.

There is also a new codec being developed by the Alliance for Open Media named AV1. The Alliance includes companies from the major hardware, software and service companies today with the goal of creating the next generation royalty-free codec. AV1 is expected to be finalized late 2017 and is expected to supersede VP9.

2.3 CODEC COMPARISON

These new codecs provide quality improvements and data savings, but the meaningful effect on mobile operators depends on the goal of the service. A service can decide to decrease the bitrate per resolution, and increase the visual quality per resolution while using the same bitrate or a combination of the two.

Netflix recently started encoding its download content using VP9 and found that it can achieve up to a 36 percent savings as shown in Figure 5. What this reduction means is that Netflix can stream higher quality video for the same bitrate.[7] While Netflix has only reported VP9 for downloads, testing with a device released in September 2017 shows that it's streaming both VP9 and H.265 content.

[7] https://medium.com/netflix-techblog/more-efficient-mobile-encodes-for-netflix-downloads-625d7b082909.

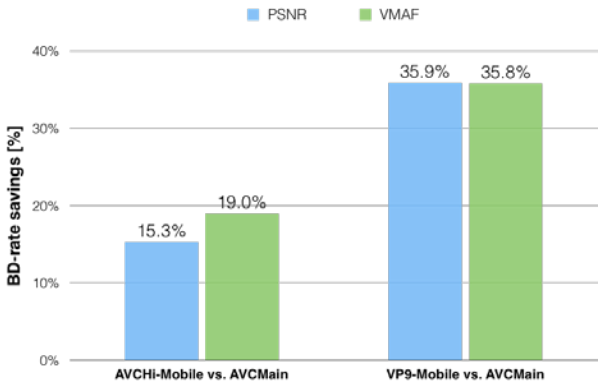

Figure 5: VP9 Data Savings.

Testing across U.S. mobile operators shows that Netflix defaults to a 650 kbps encoded video stream over mobile networks. Traditionally this has translated to 384P, but with the new codecs and their per-title encoding optimization, 540P for VP9 videos and 480P for HEVC HDR videos are currently available.[8] This done while maintaining 650 kbps default bitrate for mobile devices.

To understand how this is done for another major service, we analyzed the encoded and manifest bitrates for the top 10,000K trending YouTube videos over 90 days.[9] The encoded bitrate of a video is the average bitrate of the entire video. In other words, it is the total size divided by the duration of the entire video. The encoded bitrate translates to the volume used per resolution.

The manifest bitrate is the average bitrate of the largest video chunk of the video. As noted previously, adaptive bitrate streams are broken into chunks of video. The manifest bitrate represents the bitrate of the largest of these chunks per resolution. This can be represented as the full bitrate of the largest chunk or as a percentage of bitrate for this chunk. The manifest bitrate is used by the video client to determine which video chunk to request based on current throughput levels.

Table 1 shows this study's results. The table includes the resolution, frame rate, encoded bitrate and manifest bitrate for the video. It does not include the audio bitrate, which typically adds 128 kbps per video resolution.

There are a few important things to note about the information provided here. First, the analysis was based on the top videos on Google trends. This may have skewed the results by preferring one type of video or content creator over others, however, this is a valid data point because the videos analyzed were the top trending videos each day. Second, the analysis was based entirely on the information provided in the manifest file for each video. Quality comparisons were not performed, so there is no opinion as to whether YouTube has preferred a higher quality video over data savings per resolution.

[8] https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2.
[9] https://trends.google.com/trends/hotvideos/hotItems.

Table 1: Encoded vs. Manifest Bit Rate Comparison Between VP9 and H.264.

	H.264		VP9	
Resolution	Encoded (Mbps)	Manifest (Mbps)	Encoded (Mbps)	Manifest (Mbps)
240P	0.22	0.26	0.18	0.25
360P	0.36	0.57	0.34	0.47
480P	0.69	1.07	0.59	0.85
720P	1.30	2.07	1.18	1.68
720P (60fps)	2.51	3.40	2.14	2.94
1080P	2.50	3.77	2.20	3.01
1080P(60fps)	4.54	5.74	3.85	5.01
1440P	5.91	8.43	5.90	8.63
2160P	13.90	19.79	14.72	19.55

From this analysis, it was learned that the average manifest and encoded bitrates show up to a 21 percent and 18 percent reduction for VP9 videos, respectively. It's interesting that there is such a large range of savings and that the savings is lost for higher bitrate videos.

2.4 CONGESTION CONTROL ALGORITHM ENHANCEMENTS

The final form of service optimization we will cover is related to congestion control algorithm enhancements. Transmission Control Protocol (TCP) is a connection-oriented protocol that enables reliable content delivery between two nodes. TCP, and now Google's QUIC,[10] utilize a congestion-control algorithm to determine how quickly a connection ramps up, how severely it reacts to loss (or perceived congestion) and how quickly it recovers from loss.

Because mobile network quality varies by geography and time, and can vary greatly as a user moves through the mobile network, it can be difficult for congestion control algorithms to accurately estimate available bandwidth. This often leads to an under- or overestimation of available resources, which results in inefficient delivery and suboptimal user experience. TCP split proxies or other middleboxes are often deployed in mobile networks to help overcome the challenges that traditional loss-based TCP congestion control algorithms have in wireless.

Google recently released a new congestion control algorithm into the Linux kernel named Bottleneck Bandwidth and Round-Trip Propagation Time (BBR), which aims to overcome the challenges of traditional loss-based congestion-control algorithms. It does this by taking a new approach to detecting congestion along the connection's path.

Google's blog says: "For a given network connection, it [BBR] uses recent measurements of the network's delivery rate and round-trip time to build an explicit model that includes both the maximum recent bandwidth available to that connection, and its minimum recent round-trip delay. BBR then uses this model to control both how fast it sends data and the maximum amount of data it's willing to allow in the network at any time."[11]

[10] https://tools.ietf.org/html/draft-tsvwg-quic-protocol-02.

[11] https://cloudplatform.googleblog.com/2017/07/TCP-BBR-congestion-control-comes-to-GCP-your-Internet-just-got-faster.html.

Put another way, BBR uses bandwidth and RTT measurements to model the network with the goal of maximizing bandwidth while minimizing delays. Because it isn't a loss-based algorithm, it can achieve maximize bandwidth even with non-congestion loss rates up to 15 percent.[12]

To show how this translates to user experience, Google's blog cites the following improvements achieved when it switched YouTube from Cubic (the default congestion control algorithm in the Linux kernel) to BBR:

- 4 percent higher network throughput
- 33 percent reduction in round-trip time
- 11 percent higher mean-time-between-rebuffers

To understand how this works on an LTE network, experiments with various RAN and backhaul configurations were performed; it was found that BBR flows performed better than Cubic flows in many scenarios. It is recommended that a more comprehensive study on what the adoption of BBR will mean for mobile networks be performed.

The optimizations and improvements reviewed address the overall tonnage and delivery improvements by the industry and services; however, it does not necessarily address the coverage or all the congestion challenges mobile operators face. The next section covers how mobile operators can use new features to more effectively manage video traffic and user experiences under various network conditions.

3. MANAGING VIDEO BASED ON NETWORK LOAD

Next we cover four aspects from the network perspective. Some of these require partnerships between the video service and mobile operator, but they would all benefit from both service and operator working together to maximize each technology's potential.

3.1 RADIO CONGESTION AWARE FUNCTION (RCAF)

The Radio Congestion Aware Function (RCAF) was introduced in 3GPP Release 13 to report Radio Access Network (RAN) User Plane Congestion Information (RUCI) status to other network elements. RCAF does this by collecting RAN information from the Operation Administration and Management (OAM) interface and user information from the Mobility Management Entity (MME) over the Nq interface. It then notifies either the Policy Controls and Charging Rules Function (PCRF) or Service Capability Exposure Function (SCEF) which users are in a congested state based on a set of reporting rules. RCAF can report RUCI to the PCRF via the Np interface for account policy decisions or the SCEF via the Ns interface to pass the information on to partners or third parties.

For mobile video, RCAF can be used to dynamically optimize content based on the congestion level. In the case of PCRF policy decisions, the PCRF can trigger the Policy and Charging Enforcement Function (PCEF) or some other internal node to optimize content while the subscriber is in the congested state. In the case of the SCEF, the mobile operator can notify third-party service providers to adjust delivery policy to the mobile device.

It is also worthwhile to extend the functionality of the RCAF-SCEF-AS architecture so that user level congestion information can be provided to external parties, such as the information granularity available to the PCRF over the Np interface. This is, however, outside of the scope of existing specifications and may

[12] https://www.ietf.org/proceedings/98/slides/slides-98-iccrg-an-update-on-bbr-congestion-control-00.pdf.

involve individual development and customization on the RCAF and SCEF platforms. Because the RCAF already has a specified interface, Nq, via the MME, it follows that the user-level information can already be derived by the RCAF. For this extended functionality to be realized, additional Attribute Value Pairs (AVPs) should be integrated over the Ns interface so the RUCI includes the user level information. The SCEF will then have to translate the user level information (e.g., IMSI) to a transcoded ID so it's not exposed to the external party. Of course, the operator and the external party must agree on the transcoded ID format, encryption and decryption.

Potential use cases include:[13]

1) Adjusting content and delivery based on network conditions. This informs the third-party service provider of congestion status or general load level.

2) Enabling background transfers and pre-fetching of content. This informs the third-party service provider about the different recommended time windows for data transfer to specific UEs in a geographical area, the maximum bitrate that can be handled during these different time windows and the applicable charging rates.

Because the mobile operator may not want to expose congestion information externally, special consideration will need to be taken when exposing this information.

3.2 MOBILE THROUGHPUT GUIDANCE

Mobile Edge Computing (MEC) enables cloud computing at the edge of the cellular network. One interesting use case for MEC to address some challenges with mobile video is the Internet Engineering Task Force (IETF) draft Mobile Throughput Guidance proposal.[14]

As discussed in previous sections, congestion-control algorithms often have challenges estimating the available bandwidth on mobile networks. To overcome these challenges, the Mobile Throughput Guidance architecture was developed to pass throughput guidance from the mobile network to the serving node. This node can be the application server or a middlebox residing inside the mobile network. As outlined in Figure 6, throughput guidance does this as follows:[15]

1. MEC inserts throughput guidance in the uplink Transmission Control Protocol (TCP) options.

2. TCP Server uses this information to assist in TCP congestion control decisions and to ensure that the application level coding matches the estimated capacity at the radio downlink.

3. A trustful relationship is established between the Throughput Guidance (TG) provider and the TCP server

[13] 3GPP TS 23.708 SA#68.
[14] https://tools.ietf.org/html/draft-flinck-mobile-throughput-guidance-04.
[15] https://www.ietf.org/proceedings/92/slides/slides-92-tsvwg-12.pdf.

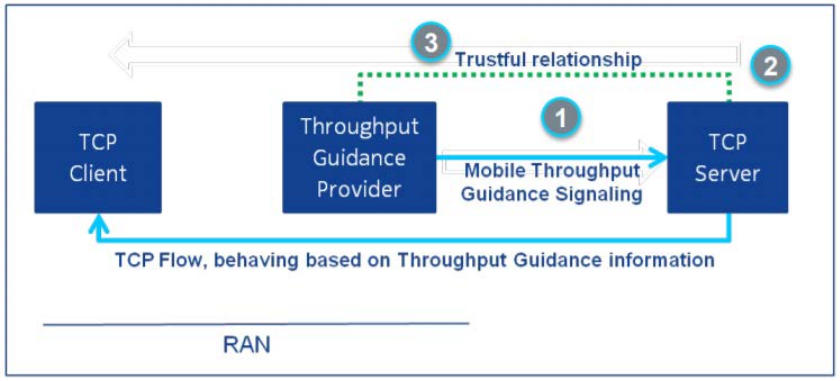

Figure 6: Throughput Guidance Provider Signaling Flow.

By exposing this information to the TCP server, the server and application can more accurately determine available bandwidth and adjust the delivery rate as necessary. A trial run by Nokia, Google and Vodafone showed a 30 percent to 60 percent improvement in network metrics and a 10 percent to 15 percent improvement in application metrics[16] when using this method.

These improvements highlight the benefits of collaboration between mobile operators and application services; however, the flow of information is currently one direction in this standard. It would be beneficial to add a bi-directional flow of information where the application server notifies the mobile operator in-band of degraded user experience. This information could be passed through the same mechanism that throughput guidance is using. This would allow the mobile operator to identify and address trouble spots in near real-time similar to how the application server adjusts delivery in real time.

3.3 SELF ORGANIZING NETWORK (SON)

Many operators have deployed Self Organizing Network (SON) technologies to better optimize the RAN. Such SON applications help balance load across cells and make best use of the available RAN capacity. However, today's SON technologies are not application-aware. They instead make their load-balancing decisions by examining radio parameters and bandwidth utilization information.

Adaptive Bitrate (ABR) video poses an additional challenge because its bandwidth adapts to available bandwidth. So although two neighboring cells may have similar utilization, in one cell the ABR videos may be adapting down to available bandwidth with subscribers having poor quality of experience (QoE) with video being viewed. Meanwhile in the other cell, the videos may be adapting up to available bandwidth, with subscribers experiencing excellent QoE for their videos. From the service provider point of view, it would be better if all users received good quality rather than some experiencing poor and others excellent quality.

Enhancing SON applications to be video-aware can help improve the described situation. With this enhanced approach, the SON applications would be aware of video sessions in the different cells and the QoE achieved by them. The SON application can then take video QoE in each cell, besides the utilization and RF parameters on each cell, to better balance load across the cells. This will ensure better use of RAN resources while maximizing aggregate video QoE across the RAN. Given that more and more of the mobile network traffic is video, video-aware SON is an important tool for operators to have in their arsenal.

[16] https://www.ietf.org/proceedings/92/slides/slides-92-tsvwg-12.pdf.

Video-aware-SON would be even more effective when content providers and mobile operators cooperate so that the video QoE information is made available by content providers. That way, mobile operators can optimize the network to improve video QoE.

3.4 5G QOS PARAMETERS

There are several QoS parameters in 5G, namely the 5G QoS ID (5QI) and Allocation and Retention Priority (ARP), which are used for a Non-Guaranteed Bitrate Flow. In case of a Guaranteed Bitrate Flow, there are also the addition of the Guaranteed Flow Bitrate and Max Flow Bitrate (GFBR and MFBR) for both uplink and downlink, as well as the Notification Control.

In LTE, QoS is applied on a bearer level. In 5G, it's applied on a flow level. The QoS flow is the finest granularity of QoS differentiation in a PDU session. A QoS Flow ID (QFI) is used to identify a QoS Flow in 5G. User plane traffic with the same QFI within a PDU session receives the same traffic forwarding treatment (e.g., scheduling, admission threshold).[17]

The 5QI is a reference that 5G uses to map the QoS Flow to forwarding treatment parameters. These parameters are called 5G QoS Characteristics, which consists of Resource Type, Priority Level, Packet Delay Budget, Packet Error Rate and Averaging Window. The 5G system can either use the standardized 5QI values or non-standardized values. In the latter's case, the QoS characteristics should be signaled from the core to the access network as part of the QoS profile.

ARP (Allocation and Retention Priority) is composed of three individual parameters: Priority Level, Pre-Emption Capability and Pre-Emption Vulnerability. The ARP is used to determine whether a new QoS flow may be allocated or denied a resource in case of network resource limitations. Furthermore, it is also used to determine whether an existing QoS flow may be pre-empted or not to accept and allocate a resource for the new QoS flow. This is resolved via the Pre-Emption parameters. The Pre-Emption Capability indicates whether a flow may get resources from other flows with lower priority. On the other hand, the Pre-Emption Vulnerability indicates whether an existing flow may lose the resource assigned to it to give way for a new flow with a higher priority.

The Guaranteed Flow Bitrate (GFBR) and the Max Flow Bitrate (MFBR) indicate and limit the bitrate that may be provided by a Guaranteed Bitrate QoS flow, respectively.

Notification control is used as an indication whether notifications are used between the RAN and the core network. If this parameter is enabled, and in the case when the RAN determines that the requirement to support the GFBR for a QoS flow can no longer be met, the RAN signals a notification to the core. Moreover, once the conditions improve and the GFBR is met, the RAN sends a new notification to inform the core network.

The idea of Reflective QoS is also introduced in 5G so the network can implicitly signal to the mobile the QoS rules that the mobile needs for classification and marking of Uplink (UL) packets. This is done by sending a Reflective QoS Indication (RQI bit being sent) to the mobile. When the mobile receives this indication and if it supports Reflective QoS, then the mobile shall derive the QoS rules based on the received downlink traffic.

[17] 3GPP TS 23.501 section 5.7.1.1.

3.5 VIDEO DELIVERY OPTIMIZATION IN 5G

The 5G QoS framework is well-suited to provide an efficient way of delivering data packets to and from the network. It enables an adaptive and real-time allocation and management of network resources to meet traffic demands. The finest granularity of data traffic management can be achieved via the 5G QoS framework because the 5G QoS parameters are intrinsic not only to the individual network functions in the core network, but also to the RAN and mobile device. With this, end-to-end optimized traffic delivery is possible with the 5G QoS. Because the PCF is part of the 5G core network and is also involved in the QoS framework, it follows that policy-based video traffic optimization is also possible.

Policies and QoS settings in 5G can be tailored to address specific needs of different traffic profiles, such as live video streaming, video download, cached video streaming and VR. Furthermore, the policies can be configured based on subscription profiles where higher priority is allocated for premium services. The different QoS parameters can also be fine-tuned individually, so the resulting QoS profile is well suited to deliver a specific traffic profile. For example, priority level in the Allocation and Retention Priority (ARP) parameter is set to a relatively high value (less than 9) with Pre-Emption Capability enabled, and Resource Type in 5QI set to GBR with a priority level also set to a relatively high value (40 or less) for high-paying subscribers who opt-in for premium services. In contrast, the same premium service can be made available to non-premium subscribers but instead with an ARP set to a lower priority, with Pre-Emption Vulnerability enabled instead. In addition, the Guaranteed Flow Bitrate (GFBR) and the Max Flow Bitrate (MFBR) can also be set to a lower value for this use-case.

The 5G core network has a dedicated network function that handles all user plane interaction between the access and the data networks. This is the User Plane Function, whose Service Data Flow templates are used to classify individual QoS flows. That means the network can also be designed to be flexible so that different QoS profiles are applied to different traffic flows, such as YouTube and Netflix having distinct QoS profiles.

A wide range of possibilities in optimizing video delivery is feasible via the 5G QoS framework. As of this writing, there are only 15 standardized 5QI values. With new video-centric services in the future, it's likely that the standardized 5QI values will be extended. On the other hand, mobile operators can always resort to non-standard QoS characteristics to support new services.

To summarize, there are new functions within the mobile network to optimize video content, but this requires the ability to correctly classify the traffic. This will be increasingly difficult to do without feedback from application services, now that the majority of traffic has moved to encrypted protocols. While the mobile operator has new tools to apply policies themselves based on network load, it is recommended to have an inclusive ecosystem with a bi-directional flow of information between mobile operators and application services to ensure the best possible user experience. This is important given the scarce availability of radio resources.

4. GEOFENCING FOR LICENSED CONTENT DELIVERY

4.1 GEOFENCING BACKGROUND

The term geofencing is used in various contexts, all of which involve a virtual perimeter overlaid on a geographic area (the geofence) and mobile devices. It involves determining the current geographical location of a mobile, comparing the current location with respect to a preset geofence and, based on a predetermined set of triggering conditions, taking actions or preventing actions from being taken. The

motivation for creating a geofence generally derives from one of two categories of uses cases: proximity to a point of interest or location of the mobile device with respect to the geofence perimeter. The latter is of primary interest for this paper.

For a geofence created for proximity, the coordinates of a point of interest (POI), typically expressed as latitude and longitude, form the center of a circle (other geometrical shapes are also possible) of a certain radius. Service providers take preset actions, typically by sending notifications, when a mobile device is determined to be in proximity—i.e. inside the geofence—of the POI or POIs.

For licensed content delivery, however, the focus shifts to the area covered by the geofence perimeter rather than a specific POI. When a mobile device enters or exits the geofence, it is considered a breach of the geofence perimeter, and a corresponding notification is generated by the geofence software. The geofence software can be run on the mobile device or on a server in the network that interfaces with a device-resident app requesting geofencing service. To detect geofence breaches, a mobile device's location needs to be known relative to the geofence. This requires periodic tracking of the mobile's location. The presence of the device inside or outside the geofence is the primary determinant for whether licensed content is allowed to be streamed to the device.

A geofence is defined both by its shape[18] and its geographical location. While a geofence can assume any arbitrary shape, for practical purposes the most commonly found geofence shapes are circles, ellipsoids or polygons. Figure 7 is an example of a circular geofence.

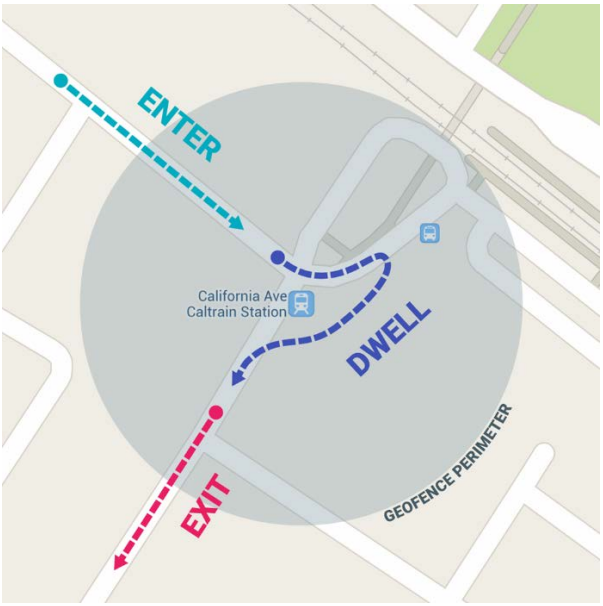

Figure 7: Geofence Perimeter.[19]

The key elements of a geofence for licensed content delivery involves:

- A geofence definition
- A mobile device to receive the licensed content
- A mechanism for locating the mobile device relative to the geofence

[18] While three dimensional geofences are possible, we'll limit ourselves to two dimensional geofences for the sake of simplicity.
[19] https://developer.android.com/training/location/geofencing.html

- A content delivery network that implements controls based on a mobile device's location relative to the geofence
- Other factors unrelated to device location—such as subscriber profile, date and time of request, content type—necessary for geofence implementation

While the exact details of geofence implementation for licensed content delivery are out of scope for this white paper, it is nonetheless instructive to look at some general features of a geofence that apply in this context:

1. The area of interest covered by geofence needs to be defined by the content service provider based on applicable license restrictions. The definition can be as simple as picking a single lat/long as the center of the geofenced area along with a radius for the size, or as complex as a set of polygons (i.e., sets of lat/longs defining the vertices of the polygon). The geofence definition also contains several control parameters such as a star/stop data option, dwell time and/or periodicity of device location. A geofence may even consist of a set of individual geofences bundled together to form a single, larger geofence.
2. There are any number of technical solutions available for determining the location of a mobile device. These include satellite positioning (GNSS), terrestrial positioning (triangulation of cell tower or Wi-Fi signals), modern sensor-based positioning (e.g. dead, reckoning based on accelerometer and/or gyroscope measurements) or combinations thereof. In short, modern-day mobile devices are location aware. However, location information sourced from a device itself suffers one fundamental flaw: It cannot be trusted. The primary source of location information for licensed content delivery needs to be the serving mobile network, not the device requesting video content. We discuss this issue further in a later section.
3. Geofence "breach event" detection functionality determines the device's location relative to the geofence. Once the geofence application has been implemented and initialized with control parameters, it will be able to track a mobile with periodic location determination. The periodicity is set as a geofence parameter to suit the specifics of the use case
4. Once a geofence breach event has been detected, content delivery can be handled appropriately (i.e., started, stopped, resumed, or paused). This is achieved by proper policy-based controls established between the licensed content provider (e.g., via a video content server) and the mobile network service provider.

In an ideal world, a mobile device would be able to participate in the detection of geofence breach events, resulting in high accuracy, low delay and high confidence outputs while consuming little power. However, trade-offs need to be made. For instance, the need for low power usage may imply limited geofence breach responsiveness (i.e., the breach may not be detected instantaneously but only after a certain delay) and some lowering of geofence breach confidence (i.e., the probability that a detected geofence breach occurred). It is up to the video content provider to engage with the mobile network provider and app developer (and potentially also the device/chip manufacturers) to determine the optimum usage envelopes for specific licensed content applications such as real-time streaming or content download with later viewing. With content downloads, for instance, geofence breach responsiveness may not be of great importance and can be traded off against power consumption. However, breach detection confidence must not be jeopardized so as not to violate content licensing terms.

4.2 LICENSED CONTENT DELIVERY

In the current context, the applicable triggering condition is the binding licensing agreement that limits distribution/consumption of video content beyond preset geographical boundaries (the geofences). Content owners are concerned about geofiltering for any business deals with content distributors when the content license is restricted to certain geographical boundaries. Because nearly all commercially produced video content is subject to some form of licensing restrictions, the need for geofiltering is nearly universal for this category of media. Following is a brief review of commercial distribution channels for video content.

In its "17th Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming" (May, 2016), the FCC categorizes the entities providing broadcast video services into three groups:[20] traditional broadcast television stations, Online Video Distributors (OVDs) and Multichannel Video Programming Distributors (MVPDs).

Broadcast TV

Traditional broadcast TV service providers are subject to broadcast license agreements under FCC jurisdiction. A broadcast license grants the permission to use a portion of spectrum for broadcasting content. Importantly, in this context, the license also limits distribution of content to a pre-defined geographical area. Licensed commercial operators that typically provide television, radio, and other two-way communications services, receive FCC assignments to a portion of spectrum (single or multiple band) for their operation.

Traditional broadcast TV stations must adhere to the geographical constraints for content distribution embodied in their license agreements. However, although full-power television stations have transitioned to digital transmission and have the capability to offer additional multicast linear digital channels, they still offer far fewer programs and channels than are available from MVPDs, and do not provide subscription-based services. Broadcast television stations form a distinct category and are not the focus of this paper.

OVD

Online Video Distributors (OVD) is an entity that provides video programming over the Internet (over any IP-based transmission path) where the actual transmission path is provided by an entity other than the OVD itself.

An MVPD's service area is defined by the provider's dedicated (either owned and/or leased) distribution infrastructure. However, a broadcast TV station's service area is defined by its signal coverage area and designated marketing area (DMA). An OVD's geographic service area can potentially cover any and all regions capable of receiving high-speed Internet service. OVDs rely exclusively on IP-based transport, including both public Internet and private infrastructure, to deliver their content to the consumers. OVD users can access online video via Internet-enabled devices, such as computers, smartphones, tablets, gaming consoles and TV sets, as long as the device has broadband connectivity.

OVDs typically handle multiple categories of video content. In addition to professionally produced commercial content, OVDs may also handle other categories of videos such as consumer/user-generated videos that was produced with professional-grade equipment and publicly available content created by end users.

[20] *Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming [Seventeenth Report; MB Docket No. 15-158; DA 16-510]* FCC. May 6, 2016.

Subscription OVDs (ad-free service) charge users either monthly or annual fees for the right to stream content. Subscription OVDs may negotiate with both content generators (e.g., studios) and content distributors that own fixed distribution infrastructure to acquire distribution rights for movies and television series. Typically, subscription OVDs negotiate for older TV shows and film libraries of movie studios. A few may also provide quick access (next day) for some time-sensitive TV content.

It is important to note that depending on the distribution window, licensing agreements may be exclusive to OVDs or non-exclusive. Movie and television studios are cautious in licensing content to subscription services to minimize impact on revenues from in-theater or Blu-ray DVD sales, typically resulting in a 1-3-month delay in availability for newly released movies. Also, geographical restrictions on the distribution of content applies to OVDs, just as they apply to the MVPD and broadcast TV categories.

While the FCC does not define the term Over-The-Top (OTT), it is typically used to refer to audio, video and other media transmitted via the Internet as a standalone product without the direct involvement/control of an operator of the underlying cellular, cable or direct-broadcast satellite systems used in distributing the content. OVD can be looked at as a special case of OTT.

Geographical restrictions on the distribution of content apply to OVDs just as they apply to MVPD and broadcast TV. However, this paper doesn't discuss how licensing restrictions may apply to content sourced by OVDs or how such restrictions may be implemented by OVDs. OVDs are noted as an important category of video content providers deserving of further attention.

MVPD

The U.S. government defines Multichannel Video Programming Distributors (MVPDs) as an entity such as, but not limited to, a cable operator, a multichannel multipoint distribution service, a direct broadcast satellite service or a TV receive-only satellite program distributor. These providers sell multiple channels of video programming. Major MVPDs offer hundreds of linear television channels (programs on specific channels at specific times) and thousands of non-linear video-on-demand (VOD) channels. MVPDs are also eligible to retransmit.

An MVPD is an entity that provides subscription-based service for multiple channels of video programming that includes VOD and pay-per-view (PPV) programs. MVPDs frequently bundle other services, such as Internet and phone, as core elements of their business models. Video programs are typically available on multiple device formats such desktops, laptops, Internet-connected TVs, smartphones or tablets.

There are three major categories of MVPD service providers: cable, Direct Broadcast Satellite (DBS) and telephone operators. Based on a late 2014 FCC report,[7] cable accounted for 52.8 percent of MVPD subscribers, DBS accounted for 33.8 percent and telephone operators accounted for 13.0 percent. Table 2 provides a breakdown by various categories for 2013 and 2014.

Table 2: Homes Passed by MVPDs (in millions).

Type	Year-End 2013	Year-End 2014
CABLE	**131.6**	**132.2**
- Comcast	53.8	54.7
- Time Warner	29.9	30.5
- Charter	12.8	12.9
- Cox	10.4	10.5
- Cablevision	5.0	5.0
DBS	**132.9**	**133.5**
- DIRECTTV	132.9	133.5
- DISH	132.9	133.5

Tel Operator	**48.5**	51.1
- AT&T U-Verse	27.0	**28.0**
- Verizon FiOS	18.6	19.8
- Century Link	2.1	2.4
- Consolidated Comm	0.5	0.6
- Cincinnati Bell	0.3	0.3

The data in Table 2 was taken from a 2016 FCC report and are for illustrative purposes only. Recent developments, such as the merger of DIRECTV with AT&T, are not reflected.

For all categories of video content distributors mentioned above, ownership of exclusive territorial rights to content offered may differ between geographical regions. The distributors are bound by the licensing terms and conditions for content that disallow access for users outside of their designated region.

For example, HBO is available only to U.S. residents. The reason for this restriction is that the parent company, Time Warner, has licensed exclusive rights to HBO content to various regional distributors (e.g., HBO is licensed to Bell Media in Canada) that may offer their own, similar and region-specific service that competes with the HBO business model. Similarly, content available on subscription VOD services (e.g., Amazon, Netflix) may also vary widely, or be entirely blocked, from region to region.

The licensing restrictions are imposed for various reasons, both commercial (revenue generation) and regulatory (e.g., promotion of competitiveness, adherence to applicable local tax/media/Internet regulations). For this paper, the underlying reasons for imposing licensing restrictions are not relevant.

Key questions for geofencing are: how can providers of video services (streaming or otherwise) establish with high confidence the location of a device running the service's video app or viewing the service via a browser? From the licensed video service provider's perspective, the service areas relevant to licensing restrictions are defined by:

- National boundaries consisting of one country or set of countries
- Regional boundaries inside a country or set of contiguous countries. Sometimes regional boundaries for authorized service areas may be defined by a collection of postal codes

The video content licensees are required to enforce various restrictions on distribution. For example, sports/events blackouts must be enforced in areas close to arenas hosting the live event, or outside a local broadcast channel's Nielsen DMA, or in states served by a regional sports network (e.g., Big 10 Network). A tool for implementing such control is geofencing.

Figure 8 is the Nielsen DMA 2012-2013 map for the U.S. It provides a general idea of the size of geofencing boundaries relevant in this context. We note that area sizes involved are large (or very large) compared to conventional use of geofencing technology. As explained later, the large area size is the key factor for selection of a positioning mechanism for determining device location. High-accuracy, device-based location determination for large geofence areas is not only an unnecessary overkill, it also opens the door for fraud. The positioning mechanism must combine sufficient accuracy with a high level of trust, which is a key requirement in this context.

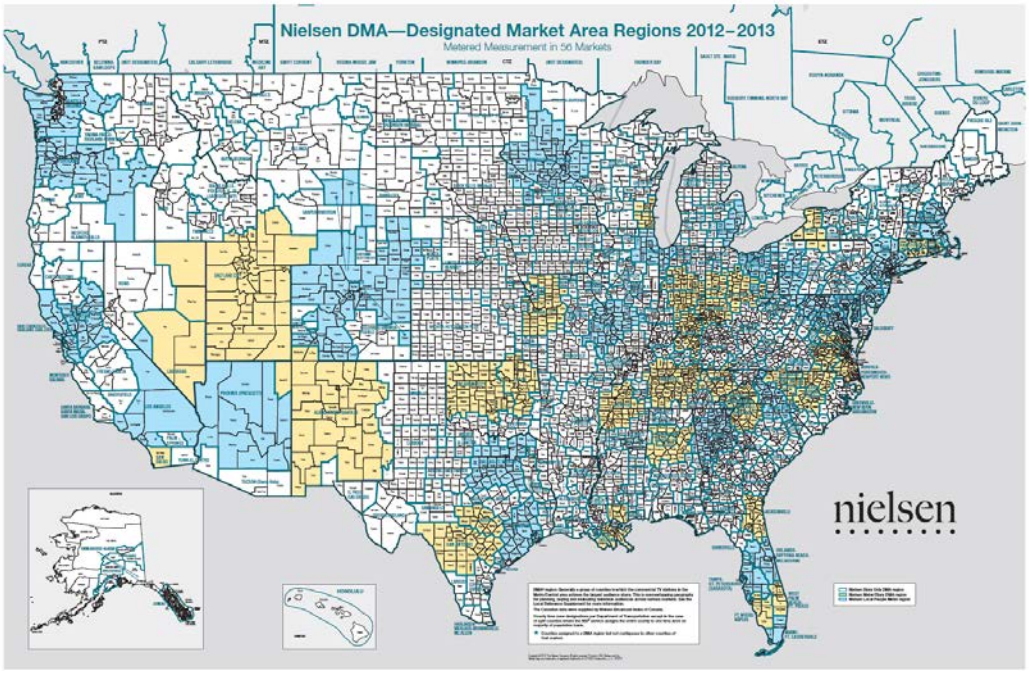

Figure 8: Nielsen DMA 2012-2013.

4.3 ACCESS VIA MOBILE NETWORKS

Multichannel Video Programming Distributors (MVPDs) use Content Delivery Networks (CDNs) to deliver video to their subscribers who, if mobile, receive those services over cellular or Wi-Fi. Figure 9 is a simplified diagram of the mobile video delivery architecture for distribution of video content, both live and stored. The CDN is not necessarily physically distinct from the mobility network but is, at least in part, a logical overlay that reuses much of the mobility network's physical backbone and access infrastructure. From the point of view of logical control, the CDN does function as an autonomous entity controlling all aspects of video delivery such as subscription management, storage, live feeds, optimum bit rate selection, encryption and security. It is assumed that this is done in close coordination with the underlying mobility network. Among the CDN functions mentioned, the focus of this paper and a primary requirement for the CDN is to ensure conformance with licensing restrictions for video delivery.

Figure 9: Mobile Video Delivery Architecture.

The CDN relies on device location information to enforce geographic license restrictions. However, device location has both regulatory and business implications. In a purely business context, location is a key element necessary for implementing controls for satisfying of the service provider's other commercial needs. Examples of location-dependent business decisions include: global launch roll-out schedules for specific programming; ad campaigns targeted for specific regions; implementing access restrictions by date, domain, geography, video player hardware and firmware, and IP address range;j and non-approved carriers.

4.4 IMPLEMENTING A GEOFENCE FOR MOBILE NETWORKS

For the vast majority of the use cases, video content is consumed over three radio access technologies: cellular, satellite and Wi-Fi. Positioning accuracy requirements vary widely based on the technology used. At a high level, the basic accuracy requirement remains the same for all delivery media types, however, it's possible that location accuracies will vary by access types.

Geofiltering for satellite-based access is not expected to be accurate. Some overspill is not only possible but inevitable. For terrestrial broadcast TV, some level of overspill is possible and acceptable, if it is limited. Even for mobile networks, with their inherent device location capabilities, occasional deviation from licensed delivery constraints may occur. However, the fact that direct-to-home or digital terrestrial TV and even mobile networks may tolerate some overspill should be used as a reason to make geofiltering as accurate as reasonably possible without compromising the required level of trust.

In this section, the basic steps involved in creating a geofence for licensed video content delivery over mobile networks are described at a high level.

1. The video content provider needs to control when, where and to whom licensed content is delivered. As indicated in Figure 9, the content provider will typically deploy a CDN that would, not only store and deliver the video content but also, support the definition, administration and maintenance of geofences for each subscriber. Such geofences can enable a range of rate plans and user specific content delivery options. Premium plans may cover a wider geographical area

(perhaps an entire county, state or even country/countries) while basic plans may be limited in geographic reach, such as covering just subscriber's home location (e.g., home ZIP code or address).

2. As mentioned earlier, while it is possible to use device-based location and run the geofencing function on the mobile device, it is not the recommended practice. To prevent location spoofing, we recommend that the mobile network be used as the source of device location and the geofencing function be also run in a network server (either in the mobile network or in CDN). The underlying governing policies (i.e., the licensing restrictions) are hosted and executed by the server (the video app server in Figure 10) via proper settings of various parameters defining the geofence.

3. A geofence-aware video viewer app will be downloaded or pushed on the mobile device. The viewer app is configured to establish IP connectivity with the remote server following user login. The app would typically be designed to provide supporting information (e.g., login ID, device type, connection status, download speed) as requested by the remote server. The app may, depending on implementation, also support an internal API to the device location engine. This would provide the server an additional path for accessing device location when network-based location may be temporarily unavailable for any reason.

4. As depicted in the diagram, the geofencing engine runs in the remote server (video app server) that controls the device-resident video viewer app. Following the request for video content, the server initiates the geofencing logic, or adds the request to an already running (based on prior request) geofencing function. Per the rules set for the geofence, the server seeks the device location (unless recently cached), and if the licensing restrictions allow, begins content streaming or download.

5. The implicit assumption underlying this architecture is close collaboration between multiple entities. It is assumed that video content service providers will engage with mobile network operators and video app developers (and potentially also device manufacturers) to ensure that the level of service, especially accuracy and trust for device location, meets the constraints imposed by licensing. This collaboration also may include other business needs originating from subscription agreements, ad promotions, launch schedules and so on.

4.5 LOCATION ARCHITECTURE FOR MOBILE NETWORKS

Previously, the justification for the reliance on network-based location determination was explained. In this section, this approach is further explained with an overview of the LTE location architecture. How this architecture may support the case when a mobile device is roaming between home and visited networks is also outlined.

Figure 10 is a high-level diagram of LTE location architecture.

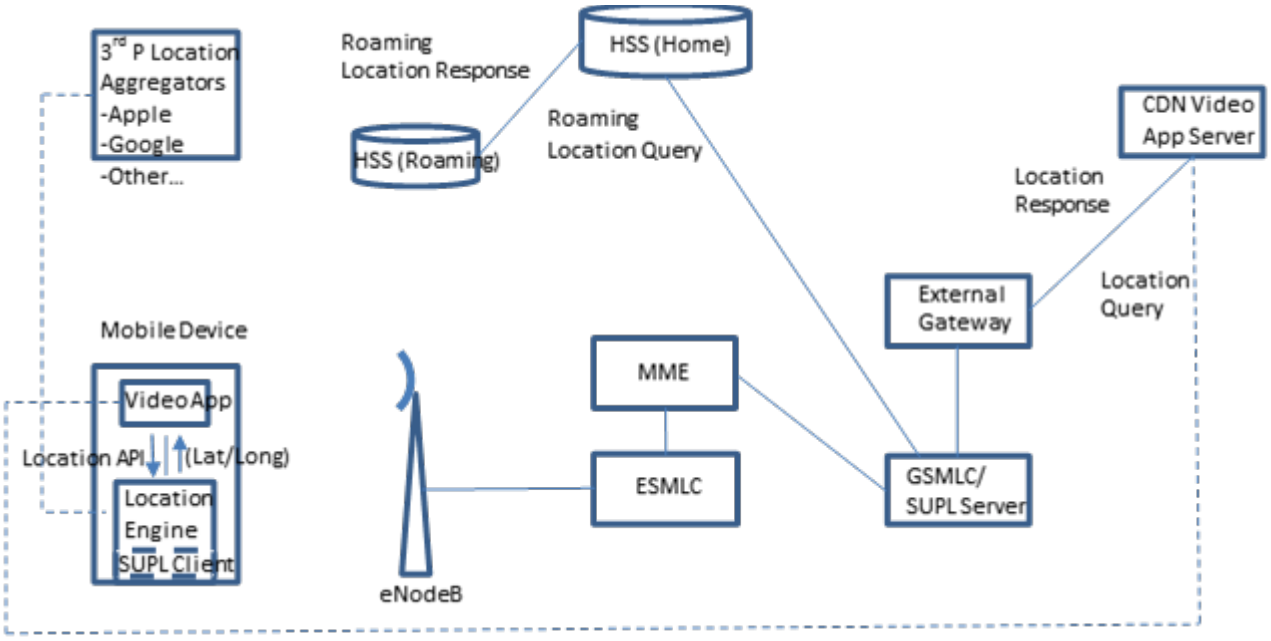

Figure 10: LTE Location Architecture.

Figure 10 shows a mobile device served by an LTE network and hosting a video viewer app either pushed or downloaded on the device. The device has an integrated location engine, a function integrated with the device Operating System (OS) for all smartphones in use today. The location engine is the source of device-based location, typically expressed as lat/long, accessible by device-resident apps over internal API interface. The location engine in turn relies on various positioning mechanisms, individually or in combination, such as Geographic Positioning Systems (GPS), user-plane location protocol (Secure User Plane Location), as well as location information coming from third-party location aggregators. The location engine responds to a location query from the app over an internal API that implements security/policy restrictions applicable to the query. Further details are not relevant for this discussion.

As previously mentioned, while device-based location determination is used by a vast majority of location-aware apps today, it is not recommended as the primary device location mechanism for licensed video delivery. The reasons include:

- Device-based location information cannot be trusted. For every device OS in use today, it is possible with software and step-by-step instructions downloaded from the Internet, to spoof device-based GPS location. Accomplishing this may not even require root-level access. A mobile device in Canada or Europe can appear to be active in Los Angeles.

- A device IP address cannot be relied upon as a reliable indicator of device location, either, because it can be easily spoofed. Additionally, there are free/low-cost VPN services that can legally hide the mobile's IP address.

- For this paper, the use case of particular interest is when a mobile device roams away from the home network. Roaming adds an additional layer of uncertainty in terms of reliability of location information coming solely from the device. However, as later explained, device-based location may be the only available near-term approach for roaming devices in the absence of standardization.

- Finally, network-based location, while not always as accurate as device-based location, provides more than sufficient accuracy for complying with licensing restrictions for video content. Most importantly network-based location is reliable—a critical requirement for licensed video delivery.

The option for device-based location, derived either from GPS or gathered from the extensive databases of third-party location aggregators engaged in continuous monitoring of device location using OTT connectivity, can be retained as a fallback option for cases where, for whatever reasons, network-based location is unavailable for a period of time. Whether to use device-based location in such situations can be left up to implementation depending on applicable CDN policies and/or the choice of geofencing parameters.

For network-based location, knowledge of the cell/sector currently serving the mobile—basic information tracked by all mobile networks—provides sufficient accuracy. This information is in turn fed back into the geofence function, along with applicable policies (assumed to be running in the background) to determine if the device can be allowed to receive, or continue receiving, the requested licensed video content.

It is recommended that the geofence function be run in a backend app server (the video app server in Figure 10), instead of the device, for security and to minimize device battery drain. The app server sends a location request using a suitable device ID (e.g., Mobile Subscriber ISDN Number (MSISDN)) to the external gateway via an API. The gateway uses the ID to query the Gateway Mobile Location Center (GMLC), which is a central network node that functions as a network clearing house for all external queries for device location. The GMLC in turn determines the currently serving mobility management entity (MME), or the visited serving network, from the Home Subscriber Server (HSS) using the device ID received. It uses this information to ask an Enhanced Serving Mobile Location Center (ESMLC) associated with the serving MME to calculate a position for the device.

Based on implementation, network-initiated and network-based device location can be calculated for a registered device with or without the device's cooperation. In either case, following determination of the geographical location (lat/long) of the serving cell/sector, GMLC applies applicable policies regarding location sharing and responds back to the external gateway. The gateway in turn responds back to the app server with the requested device location. The process is repeated with a frequency set by parameter in the geofencing software running on the server. Other geofencing parameters (e.g., duration, dwell time, periodicity) also come into play for the entire video streaming operation to proceed from start to finish. Many details, unimportant in the context of this paper, have been left out of the preceding summary description.

For network-based location the situation is more complicated when the device is roaming. The process as described above works well when the device remains inside the subscriber's home network but not when it roams onto a visited network. Currently, there is no implemented standardized mechanism for sharing the serving cell/sector ID (e.g., the enhanced cell global identifier for LTE) between the visited and home networks for a roaming device. Network signaling, except for a limited set necessary for service continuation, does not propagate across network boundaries between independent administrative domains of home and visited networks.

Assuming the geofence function is run in the mobile's home network (or in the CDN that uses home network infrastructure), the device location needs to be propagated from visited to home network. Several potential approaches for accomplishing this task are provided:

- Roaming Location Protocol (RLP), which is a protocol standard developed by Open Mobile Alliance (OMA) for sharing location information between home and visited location servers, may be implemented. RLP uses Extensible Markup Language (XML) documents for inter-location server

communication for exchanging device location. It can be used for both user-plane based location architecture, which uses the Secure User Plane Location (SUPL) protocol, or control-plane based location architecture, which uses 3GPP RAN signaling.

To the best of our knowledge, there are no known deployments of RLP in the market today. It is unlikely that the current use case would provide sufficient incentive for worldwide (or even nationwide) deployment of RLP. Without broad industry-wide deployment, RLP loses its usefulness. Therefore, this existing standards-based approach holds little promise in the foreseeable future.

- In principle, it is possible for a third-party aggregator, working with mobile operators, to develop a central clearing house of trusted device locations. This approach requires multi-party coordination between the location aggregator, mobile operators and video content providers to be workable. Large-scale implementation will be complex, from both technical and business points of view, because agreements between diverse sets of organizations would be necessary. This approach may have some regional usefulness in cases where multi-party collaboration between stakeholder companies already exists.

- Another potential approach would be to develop a mechanism, preferably standards-based, for transferring the serving cell/sector information from visited to home network. Two possible methods are explained:

    - In the first approach, the mobile device knows the ID of the cell/sector (e.g., ECGI in LTE) that is currently serving it. The device-resident video app could access the ECGI (E-UTRAN CGI – Evolved UMTS Radio Access Network Cell Global ID)—assuming it has the required permissions—from the device OS via an internal API. The ECGI of the visited network's cell/sector, retrieved from the device, could be transmitted over IP to the video app server in the home network, along with a request for video service. Based on this information, the geofence function running on the app server could determine if the request can be granted. Note that the information is retrieved from the device and, as pointed out before, device-based information lacks requisite level of trust. However, spoofing ECGI information is expected to be relatively difficult compared to spoofing raw device-based GPS location.
    - The second approach involves developing direct messaging between two administratively independent mobile networks for transfer of serving cell/sector information of roaming devices to the home network. The messaging details can be developed in an Standards Development Organization (SDO), such as 3GPP, for a standards-based solution. Policies governing such a transfer of information may complicate the messaging. However, there appears to be no inherent technical hurdles.

    Both approaches suffer from a shared problem. The serving cell/sector information coming from a visited network does not readily translate to an actual device location (lat/long) for the geofence function running in the home network. For devices in the home network, this is not a problem because the home network knows the cell towers' lat/long locations. A serving cell ECGI can be easily correlated with its actual geographical location using home network database lookup. However, this is not true for a ECGI belonging to a visited network and serving a roaming device.

A third-party aggregator could provide a solution on a global, continental, national or regional scale by creating and maintaining an updated database of cell tower locations belonging to multiple mobile operators that cover the area of interest. There are companies today that provide this type of service. There are also community supported organizations (e.g., OpenCelliD) that claim to provide this information on a global scale. However, just as indicated in the second bullet, complex business agreements would be involved for

incorporating any third-party-based solution. A limited regional solution is more conceivable in the near term.

4.6 ACCESS VIA WI-FI

For delivery of video over Wi-Fi, there are two broad categories of circumstances to consider. First, when the device has only Wi-Fi connectivity and second, when the device is active in both Wi-Fi and cellular networks simultaneously. For the second scenario, the previous discussion (LTE location) applies. This section addresses the case where the device has only Wi-Fi connectivity.

For location determination for a device with Wi-Fi-only connectivity, either, or both, of two pieces of information may be utilized: device IP address and Access Point ID (APID) of the Wi-Fi access point serving the device. The IP address of the device requesting video service, and communicated to the server in the HTTP request, could, in a perfect world, provide an indication of the mobile's current location based on publicly available data on geographical distribution of IP address pools.

However, the IP address sent from the mobile device cannot be trusted. It doesn't take much technical knowhow to spoof the device IP address today; step-by-step instructions are available online. In addition, there are perfectly legal means for hiding a mobile's IP address by subscribing to a VPN service that operates via globally dispersed proxies. For a proxied device, the IP address seen by the video server is that of the proxy, not the device itself. Most VPN services also allow their subscribers to choose among geographically dispersed proxy servers to suit their current needs. A mobile device roaming in New Zealand could select a proxy server based in Netherlands to catch live broadcast of European soccer matches and then switch to a Toronto-based proxy for hockey—both activities likely breaking content licensing agreements.

Access Point IDs (APIDs) provide a more reliable means for locating a mobile device compared to IP addresses. Typically, APIDs (based on AP Medium Access Control IDentifier (MACID)) are broadcast by Access Points (APs), along with information on channels, frequencies, encryption, etc. This information is captured by mobile devices before associating and receiving service over a Wi-Fi network. The Wi-Fi service provider maintains a database, correlating APIDs with their geographical locations, for all APs under its jurisdiction. The backend video server can use the APID, retrieved from a mobile requesting video service, to derive the device location necessary for the geofence function.

Figure 11 is example of a street map covering few cities blocks with an overlay of Wi-Fi hotspots.

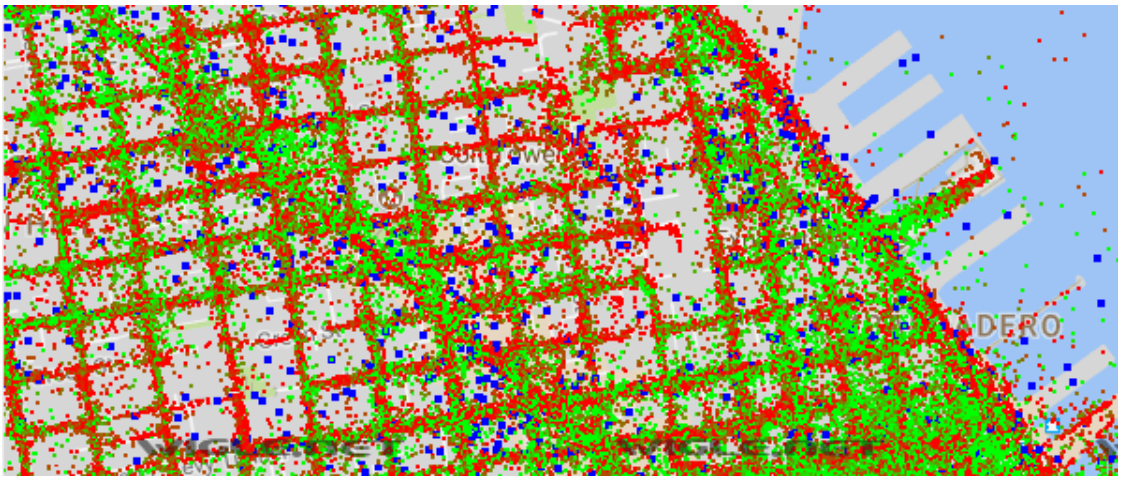

Fig 11: Wi-Fi Hotspots Mapped on City Blocks.[21]

Use of Wi-Fi APIDs shares one common requirement with ECGI: the need for an accessible central database of Wi-Fi hotspot APIDs, belonging to multiple administrative domains, correlated with their geographical locations. As before, there are many private third-party aggregators that claim to provide regional or national databases of Wi-Fi hotspots. The integrity and freshness of the information may vary. There is currently no central source that can provide a comprehensive authoritative database of hotspots spanning the entire U. S. Although such a database is currently under development in support of next-generation emergency call location, there are no current plans to allow commercial use of that platform.

For the case when video content is consumed over Wi-Fi and the device does not have connectivity with a mobile network, the only viable and trustworthy option is for the content provider to partner with a private third party (or parties) as a source of device location based on serving APID. Also, as before, with the understanding of the inherent risks involved, limited use of device-based location remains a fallback option.

4.7 GEOFENCING IN 5G

Geofences are generic in the sense that they do not depend on the underlying bearer access network used for content delivery. Apart from the fact that 5G radio access technology may enable new positioning methods specific to 5G radio (e.g., Observed Time Difference of Arrival), geofencing for licensed content delivery for 5G will be based on the same mechanisms as for previous generations of bearer access networks (i.e., 3G, 4G). At the time this white paper was written, 3GPP has not yet standardized the 5G radio access network and associated location technologies and protocols. In fact, location technology and protocols are not among the priorities of 3GPP in the early stages of 5G standardization. Given the breadth and depth of existing positioning methods and protocols, geofencing in 5G will initially be able to build on the existing standards (e.g., 3GPP Release 15 will use LPP as a positioning protocol and support the associated positioning methods such as Global Navigation Satellite System (GNSS) and Observed Time Difference of Arrival (OTDOA)). As the 5G standard evolves and matures, 5G-specific positioning methods and protocols will find their way into 5G geofencing. From an end user's perspective, this process is expected to be smooth and transparent.

[21] https://www.wigle.net.

CONCLUSION

In this paper, we highlighted two central issues involved in the delivery of video content to mobile devices: dynamic control of the delivery network including coding for optimized video delivery and geofencing solutions for compliance with licensing constraints that govern most video content distribution today.

The first part of the paper dealt with optimized video delivery from both the service and mobile operator perspective. Application services are developing new technologies to enable new services, which will place greater demands on the mobile network. There are new functions available for mobile operators to optimize content based on congestion levels, but this requires the ability to correctly classify the traffic. This will be increasingly difficult to do without feedback from application services, particularly now that most traffic has moved to encrypted protocols. While mobile operators have new tools to apply policies themselves, it is recommended that an inclusive ecosystem be established in order to ensure the best possible user experience given the scarce availability of radio resources.

The second part of the paper analyzes geofencing approaches for achieving compliance with licensing regulations when video content is consumed via mobile devices. An important fact is that in the current context geofencing involves unusually large areas compared to more conventional use of the term. The size of geofenced areas determine the location accuracy required—and consequently the optimum positioning technology for locating the mobile device.

For MVPDs, the paper focuses on delivery of services over mobile networks, and considering the large geofence areas involved, the underlying accuracy of mobile location could remain relatively coarse. In other words, the latitude/longitude coordinates associated with the serving cell/sector of the mobile network is sufficient. Highly accurate device-based location determination derived from GPS, sensors or other newer positioning technologies, are simply not necessary. Key problems of device-based position determination are the drain on battery power and, more importantly for legally binding licensed content, lack of trust in any location information originating from the device. Information coming from a mobile device, such as GPS location or device IP address, can be easily spoofed. This is true for devices running every device OS currently in use. Fortunately, network-based location is not just sufficient in its accuracy but also provide a high level of trust. In addition to using network-based device location it was also recommended that the geofence function be run in a network server instead of the device. This protects against any tampering of the geofencing software. These precautions are necessary because, for content owners operating under legally binding licensing agreements, the issue of trust ranks above most other considerations.

A high-level device location architecture for LTE offers mechanisms for deriving device location at cell/sector level (expressed as ECGI in LTE). While this is straightforward for a device that remains within its home network, the situation is more complex when the mobile roams onto a visited network. Unfortunately, the only standards-based protocol for exchanging device location between home and visited networks, RLP, has no current market deployment. It is concluded that there would be a significant role for third-party aggregators in the overall solution to the roaming location problem. Finally, it was pointed out that when all else fails, device-based location remains a fallback option as long as the inherent risk of spoofed location is acknowledged.

This white paper brings together two important strands of technology—optimized digital video delivery and geofencing for licensed distribution—that apply to the most predominant and fastest growing form of media transfer over the Internet today. The problem of video over Internet will become more complex and demanding with wider use of 3D, UHD, augmented reality and VR. The digital marketplace will continue searching for best ways to handle this explosive growth for a long time to come.

APPENDIX

A. THE STANDARDIZED 5QI TO QOS CHARACTERISTICS MAPPING. [22]

5QI Value	Resource Type	Priority Level	Packet Delay Budget	Packet Error Rate	Default Averaging Window	Example Services
1	GBR	20	100 ms	10^{-2}	TBD	Conversational Voice
2		40	150 ms	10^{-3}	TBD	Conversational Video (Live Streaming)
3		30	50 ms	10^{-3}	TBD	Real Time Gaming, V2X messages
4		50	300 ms	10^{-6}	TBD	Non-Conversational Video (Buffered Streaming)
65		7	75 ms	10^{-2}	TBD	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66		20	100 ms	10^{-2}	TBD	Non-Mission-Critical user plane Push To Talk voice
75		25	50 ms	10^{-2}	TBD	V2X messages
5	Non-GBR	10	100 ms	10^{-6}	N/A	IMS Signalling
6		60	300 ms	10^{-6}	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		70	100 ms	10^{-3}	N/A	Voice, Video (Live Streaming) Interactive Gaming
8		80	300 ms	10^{-6}	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file
9		90			N/A	sharing, progressive video, etc.)
69		5	60 ms	10^{-6}	N/A	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)
70		55	200 ms	10^{-6}		Mission Critical Data (e.g. example services are the same as QCI 6/8/9)
79		65	50 ms	10^{-2}	N/A	V2X messages

[22] 3GPP TS 23.501 table 5.7.4-1.

ACKNOWLEDGEMENTS

The mission of 5G Americas is to advocate for and foster the advancement and full capabilities of LTE wireless technology and its evolution beyond to 5G throughout the ecosystem's networks, services, applications and wirelessly connected devices in the Americas. 5G Americas' Board of Governors members include América Móvil, AT&T, Cable & Wireless, Cisco, CommScope, Entel, Ericsson, HPE, Intel, Kathrein, Mavenir, Nokia, Qualcomm, Samsung, Sprint, T-Mobile US, Inc. and Telefónica.

5G Americas would like to recognize the significant project leadership and important contributions of project co-leaders Jeffrey Smith and Salvador Mendoza of T-Mobile USA as well as Sankar Ray of AT&T along with the representatives from member companies on 5G Americas' Board of Governors who participated in the development of this white paper.