Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 22, 2016
G. Deen
NBCUniversal
L. Daigle
Thinking Cat Enterprises LLC
March 21, 2016

Glass to Glass Internet Ecosysten Introduction
draft-deen-daigle-ggie-00

Abstract

This document introduces the Glass to Glass Internet Ecosystem (GGIE). The GGIE goal is to improve how the Internet is used for all video, both amateur and professional, reflecting that the line between amature and professional video technology is increasinly blurred. As the Glass to Glass (camera lens to viewing screen) name implies GGIE's scope is from the original recording by a lens, through the steps of editing, packaging, distributed and searching, and finally viewing. GGIE is not a complete end to end architecture or solution, it is use cases and technical specifications that can serve as foundational building blocks for new Internet video innovation.

This is a companion effort to the GGIE W3C Taskforce in the W3C Web and TV Interest Group.

This document is being discussed on the ggie@ietg.org mailing list.

Table of Contents

1. Terminology

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

2. Introduction

The proliferation of users with Internet connected devices capable of capturing, and devices for watching streamed video has created what is in terms of shear bandwidth the Internet's larget use without any close second competitor. As of 2015 there are reports that youtube users upload over 500 hours of video every minute, and that during evening hours NetFlix accounts for a staggering 1/3 of Internet traffic. The number of users using the Internet for both ends of the video create-view lifecycle grows daily worldwide, and this is creating an enormous strain put on the underlying Internet infrastructure at nearly every point from the core to the edge.

While video is one of the most conceptually simple uses of the Internet, it is perhaps one of the most complex technically built from standards created by a large number of organizations and groups some dating from before the modern Internet even existed. Many critical parts of this complex ecosystem were not created with either video's particular charateristics or vaste scale of popularity in mind which has lead to both the degradation of the viewer experience, and many Internet policy issues around access to bandwidth for video and the needed infrastructure to support the continued explosion in video transport on the Internet.

Bandwidth increases at all levels of the Internet are expected to continue this is not currently, and is not expected to be the sole solution to the video scalability problem facing the Internet. To meet future expectations work is also needed to find ways of more efficiently use the network for creation, publication, and delivery of video advances in existing technology and standards.

Conceptually, people watch video when their play back device receives the encoded video data from a source, decodes it, and displays it for the viewer. However, the technical details behind making this simple concept happen are often far from simple due video viewers demand for smooth display of video frame by frame in at a constant rate without delay or skipped frames. Contiguous frames shown at a constant rate and the large size of video data combine to make distribution over Internet difficult and requiring sophisticated engineering to be done consistently well.

This document outlines the scope of the video problem for the Internet, proposes that a path forward must include foundational building blocks for video at both the application and network layers, and provides an outline of the video production lifecycle as a baseline for developing a problem statement and requirements for IETF work on the Glass to Glass Internet Ecosystem (GGIE).

3. Video is filling up the pipes

Video is without rival the top use of Internet bandwidth, and its ever growing demand for more bandwidth easily out paces the new capacity being added both globally and regionally with no let up in sight.

Continuous innovation introducing new higher resolutions, higher video quality, new distribution services, new viewing and creation devices all contribute to this every growing demand upon the Internet. The Cisco Visual Networking Index projects that by 2019 there will be nearly a million minutes of video per second transported by the Internet, a making up 80-90 of all IP traffic.

The growth in video bandwidth need is exceeding the growth in the bandwidth provisioning.

Video has been the top use of Internet bandwidth for several years and is larger than the bandwidth used by all other applications combined. This trend is not likely to ease or reverse itself as users of the Internet continue to make Internet transported video as one of their uses of the Internet, either for uploading and sharing video they creator, or as a primary sources for viewing video to a wide variety of viewing devices: computers, tablets, phones, connected televisions, game consoles, and AV receivers.

4. Video is different

Video is different than other data carried due to its extreme size of megabits per second, and gigebytes per hour of video, and when streamed for viewing its exteme sensitivity to latency and dropped packets. This makes video unique amongst all other applications using the Internet for while some have latency and packet loss sensitivities they do not have exteme data sizes, and while others may have exteme data size they do not care about latency, time to restransmit lost packets, or in some cases loss of some individual packets at all. A email user can tolerate an extra moment to retransmit dropped packets, and a web page user can tolerate a slow DNS lookup, but a video viewer sees both problems as jittery playback and as a failure of the network to meet their need. (Audio has similar challenges in terms of intolerance of delay and jitter, but the data sizes are significantly smaller).

Video data sizes continue to grow large as cameras and playback devices are able to capture and display higher quality images. Early digital video was often captured at either 320x240 pixel resolution or 640x480 standard defintion resolution. High definition or HD video at 1920x1080 became possible on some parts of the Internet

after 2011, although even in 2016 it remains unavailble or unreliable through many connections such as DSL and many mobile networks. Camera and player technology is currently expanding again to permit 4K or 3840x2160 pixel resolution reflecting a 4x data increase over HD.

Streaming is very demanding requiring consistent frame to frame playback in consistent constant time. Advanced features such as pause, fast forward, rewind, slow motion, and fine scrubbing are considered by users as standard features players that the network must support and serve to further the challenge facing the Internet.

New video abilities such as live streaming by users both one to one and one to many bring what has traditionally done professional broadcasters with dedicated broadcast infrastructure into the realm of every day users with connected smartphones using the Internet as a realtime global broadcast infrastructure.

5. Historical Approaches to supporting Video on the Internet

5.1. Video as an application

Internet video engineering began by adapting preexisting standards used for over the air broadcast (OTA) and physical media. Video encodings, such as AVI and MPEG2, originally designed for playback from local storage connected to the player where added to the data types carried by existing protocols like HTTP, and new protocols such as RTSP and HLS. Early use of the Internet for video was a copy-and-play model replacing the use of OTA broadcast and physical media to copy video between systems.

As Internet bandwidth grew sufficient to allow delivery of video data at the same rate it was being decoded, it became possible to stream video originally at very low resolutions such as 160x120 pixels (19.2 kilopixels), eventually permitting standard defintion (SD) 640x480 pixels (0.3 megapixels), and later high defintion of 1920x1080 pixels (2 gigapixels). This trend continues with some providers beginning to offer 4K or 3840x2160 pixels (8.3 gigapixels) requiring very reliable and generous Internet bandwidth end to end connection between the viewer and source.

Unlike the Web, email, and network file sharing which have been engineered and standardized in Internet focused organizations such as the W3C and IETF, video is dependent on standards developed by a very large number of groups, companies, and organizations which include the IETF, W3C but also MPEG, SMPTE, CEA, IEEE, ANSI, ISO, networking and technology companies, many others. In contrast to the extensive end to end expert knowledge and engineering done to create the Web

and email, Internet video has largely been an evolved cobbling and adaption exercise done by engineers with their focus on a few, or one, particular aspect or problem at a time, and little interaction between other parts of the Internet video ecosystem. While it is very much possible to deliver video over the Internet, this uncoordinated cobbling has resulted in many areas of inefficiency where engineering done from an end to end perspective provide the opportunity to vastly improve how video uses the Internet, which offers the hope of improving the quality of video and increasing the amount of video which can be delivered instead of relying solely on bandwidth growth to enable growth.

5.2. Video as a network problem

Network, video, and application engineers have constructed elaborate solutions for dealing with bandwidth and processing limitations, network congestion, lossy transport protocols, and the ever growing size of video data. These solutions commonly fall into one of several solution types:

1. Reducing data sizes through resolution changes, compression, and more efficient encodings

2. Downloading before playing instead of realtime streaming

3. Positioning the data close to the viewer via caches, typically on the network edge

4. Fetching of video data at a rate faster than playback

5. Transport protocols that attempt to deliver video data such that the data arrives as if it were done on a congestion free/lossless network

6. Dynamic reselection of sources and transport routes on either a realtime or frequent intervals, 10-15 seconds, using player feedback mechanisms or network telemetry

6. GGIE: Building blocks to support video through network and application

GGIE, the Glass to Glass Internet Ecosystem, is an effort to improve video's use of the Internet by examining the end to end video ecosystem from the glass lens of the camera through to the glass the screen, and to identify areas of simplifications, standardization, and reengineering to make better use of bandwidth enabling smarter network use by video creators, distributors, and viewers. GGIE is focused on how video uses the Internet, and not on how it is encoded

or compressed. Like wise GGIE does not deal with content protection. GGIE scope however does include creator and viewer privacy, content identifiction and recognition as a means to enable smarter network usage, edge caching, and discoverability.

Beyond improving the simplistic task of a viewer using the Internet to watch linear video, it is hoped, that through having a set of improved Internet video standards, the innovators can build using such standards as a foundation to create the next generation of Internet video such as multisource personalized composite experiences, interactive stories, and live personal broadcasting to name a few.

Due to the very diverse and large deployment of existing video playback devices and infrastructure, it is viewed as essential that any evolved ecosystem continues to work with the majority of the legacy deployment.

6.1. Affected IETF work areas

It is expected that significant improvement is possible in the video transport ecosystem by modest evolution and adaption of existing standards for addressing, transporting, and routing of video data flows between sources and display.

6.2. Related work: W3C GGIE Taskforce

A companion effort was begun in 2015 in the W3C Web and TV Interest Group's GGIE Taskforce. The W3C GGIE group developed a series of use-cases on discovery, search, delivery, identity, and metadata which can be found at https://www.w3.org/2011/webtv/wiki/GGIE_TF

7. Setting the stage for GGIE

This section outlines the details of the video lifecycle -- from creation to consumption -- including the key handholds for building applications and services around this complex data. The section also provides more detail about the scope and requirements of video (scale of data, realtime requirements).

Note: this document only deals with streaming video as used by movies, TV shows, news broadcasts, sports events, music concert broadcasts, product videos, personal videos, etc. It does not deal with video conferencing or WebRTC style video transport.

7.1. Media Lifecycle

The complex workflow of creating media and consuming it is decomposable into a series of distinct common phases.

7.1.1. Capture

The capture phase involves the original recording of the elements which will be edited together to make the final work. Captured media elements can be static images, images with audio, audio only, video only, or video with audio. In sophisticated capture scenarios more than one device maybe simulatneously recording.

7.1.1.1. Capture Metadata

The creation of metadata for the element, and for the final video begins at capture. Typical basic capture metadata includes Camera ID, exposure, encoder, capture time, and capture format. Some systems record GPS location data, assigned asset ids, assigned camera name, camera spatial location and orientation.

7.1.2. Store

The storage phase involves the transport and storage of captured elements data. During the capture phase, an element is typically captured into memory in the capture device and is then stored onto persistent storage such as disc, SD or memory card. Storage can involve network transport from the recording device to an external storage system using either storage over IP protocols such as iSCSI, a data transport such as FTP, or encapsulated data transport over a protocol such as HTTP.

Storage systems can range from basic disk block storage, to sophisticated media asset libraries

7.1.2.1. Storage Metadata

Storage systems add to the metadata associated with media elements. For basic block storage, a file name, file size is typical, as is a hierarchical grouping, and creation date, last-access date. For library system a identifier unique to the library is typical, a grouping by one or more attributes, a time stamp recording the addition to the library, and a last access time.

7.1.3. Edit

Editing is the phase where one or more elements are combined and modified to create the final video work. In the case of live streaming, the edit phase maybe bypassed.

7.1.4. Package

Packaging is the phase in which the work is encoded in one or more video and audio codecs. These maybe produce multiple data files, or they may be combined into a single file container. Typically it is in the packaging phase is the creation or registration of a unique work identifier for example an Entertainment Identifier from EIDR.

7.1.4.1. Package Metadata

7.1.5. Distribute

The distribute phase is publishing or sharing the packaged work to viewers. Often it is uploading it to a site such as YouTube, or Facebook for social media, or sending the packaged media to streaming sites such as Hulu.

It is common for the distribution site to repackage the video often transcoding it to codecs and bitrates chosen by the distributor as more efficient for their needs. Distribution of content expected to be widely viewed often includes prepositioning of the content on a CDN (Content Distribution Network).

Distribution involves delivery of the video data to the viewer.

7.1.5.1. Distribution Metadata

Distribution often adds or changes considerable amounts of metadata. The distributor typically assigns a Content Identifier to the work, that is unique to the distributor and their content management system (CMS). Additional actions by the distributor such as repacking and transcoding to new codecs or bitrates can require significant changes to the media metadata.

A secondary use of distribution metadata is enabling easy discovery of the content either through a library catalog, EPG (electronic program guide), or search engine. This phase often includes significant new metadata generation involving tagging the work by genre (sci-fi, drama, comedy), sub-genre (space opera, horror, fantasy), actors, director, release date, similar works, rating level (PG, PG-13), language level, etc.

7.1.6. Discover

The discover phase is the precursor to viewing of the work. It is where the viewer locates the work either through a library catalog, a playlist, an EPG, or a search. The discover phase connects interested viewers with distribution sources.

7.1.6.1. Discovery Metadata

It is typical for discovery systems to parse media metadata to use the information as part of the discovery process. Discovery systems may parse the content to extract imagery and audio as additional new metadata for the work to ease the viewers navigation of the discovery process perhaps as UI elements. The system may import externally generated new metadata about the work and associate it in its search system, such as viewer reviews, metadata cross reference indices.

7.1.7. View

The view phase encompasses the consumption of the work from the distributor. For Internet delivered video it is typical for delivery to involve a CDN to perform the actual delivery.

7.2. Video is not like other Internet data

Video is distinctly different from other Internet data. There are a number of characteristics that contribute to video's unique Internet needs. The most significant characteristics are:

1. large size of video data (Mbps to Gbps)

2. low latency demands of streamed video

3. responsiveness to trick play requests by the user (stop, fast forward, fast reverse, jump ahead, jump back)

4. multiplicity of formats and encodings/bit rates that are acceptable substitutes for one another

7.2.1. Data Sizes

Simply put compared to all other common Internet data sizes, video is huge. A still image often ranges from 100KB to 10MB. A video file can commonly range from 100MB to 50GB. Encoding and compression options permit streaming videos using bandwidth ranging from 700Kbps for extremely compressed SD video, to 1.5-3.0 Mbps for SD video, to 2.5-6.0 Mbps for HD video, and 11-30Mbps for 4K video.

Still images have 4 dimensional properties that affect their data size:

1. number of horizontal X pixels

2. number of vertical Y pixels

3. bytes per pixel

4. compression factor for the image encoding.

Video adds to this:

1. frames per second playback rate

2. visual continuity between frames (meaning users notice when frames are skipped or played out of order)

3. discontinguous jumps between frames such as skipping forward or backwards to inserting frames from other sources between contigous frames (advertisement placement)

Each video format roughly increases by x4 the data needs of the previously resolution: (1) SD is 640x480 pixels; (2) HD is 1920x1080 pixels; (3) 4K is 3840x2160 pixels.

Video, like still images, assigns a number of pixels to store color and luminance information. This currently evolving alongside resolutions after being stagnant for many years. The introduction of high dynamic range videos or HDR has changed the color gamut for video and increased the number of bits needed to carry luminance from 8 to 10 and in some formats more.

Compression is often misunderstood by viewers. Compression does not change the video resolution, SD is still 640x480 pixels, HD is still 1980x1080 pixels. What changes is the quality of the detail in each frame, and between frames. Compression algorithms work with the video images and movement to reduce data sizes through encoding of repetitive and

Video is in its simplest form a series of still images shown sequentially over time, adding an additional attribute to manage.

7.2.2. Low Latency Transport

Viewers demand that video plays back without any stutter, skips, or pauses, which translates into low latency transport for the video data.

7.2.3. Multiplicity of Acceptable Formats

One of the unique aspects of video viewing is that there can exist multiple different encodings/versions of the same video, many of which are acceptable substitutes for one another. This is a unique aspect of video viewing and differentiates video delivery from other data transports.

An email is the email, this is what enables digital signatures to operate on the email body. One composed and sent, there is only one version of the email which is the original, untouched, acceptable version of the email.

7.3. Video Transport

7.3.1. File vs Stream

There are two common ways of transporting video on the Internet: 1) File based; 2) Streaming. File based transport can use any file transport protocol with FTP and BitTorrent being two popular choices. Streaming

File based playback involves copying a file and then playing it. There are schemes which permit playing portions of the file while it progressively is copied, but these schemes involve moving the file from A->B then playing on B. FTP and BitTorrent are examples of file copy protocols.

Streaming playback is most similar to a traditional Cable or OTA viewing of a video. The video is delivered from the streaming service to the playback device in real time enabling the the playback device to receive, decode, and display the video data in real time. Communication between the player and the source enable pausing, fast forward, rewind by managing the data blocks which are sent to the player device.

8. Conclusion and Next Steps

GGIE seeks to held address this problem by establish standards based foundational building blocks that innovators can build upon creating smarter delivery and transport architectures instead of relying on raw bandwidth growth to satisfy video's growth.

Next steps will include introducing for discusion issues relevant to the IETF based on use cases developed in the W3C GGIE Taskforce, as methods for enabling networks to identify video and enable smart routing and addressing choices by edge devices.

9. Acknowledgements

10. IANA Considerations

None (yet).

11. Security Considerations

None (yet).

12. Normative References

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

Authors' Addresses

Glenn Deen
NBCUniversal

Email: rgd.ietf@gmail.com

Leslie Daigle
Thinking Cat Enterprises LLC

Email: ldaigle@thinkingcat.com