Network Working Group
Internet-Draft
Intended status: Informational
Expires: 13 November 2025

M. Joras
A. Tomar
A. Tiwari
A. Frindell
Meta Platforms, Inc.

12 May 2025

SCONE Video Optimization Requirements
draft-joras-scone-video-optimization-requirements-01

Abstract

These are the requirements for the "Video Optimization" use-case for the SCONE topic, which broadly speaking seeks to optimize video playback experience in mobile networks by cooperative communication between video content providers and the providers of network services to end users.

Discussion Venues

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the Secure Communication of Network Properties Working Group mailing list (sadcdn@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/sadcdn/.

Source for this draft and an issue tracker can be found at https://github.com/mjoras/sconepro-video-optimization-requirements.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 13 November 2025.

Table of Contents

1. Introduction

Video traffic is already 70% of all traffic on the Internet and is expected to grow to 80% by 2028. New formats like short form videos have seen tremendous growth in recent years. Both in developed and emerging markets video traffic forms 50-80% of traffic on mobile networks. These growth trends are likely to increase with new populations coming online on mobile-first markets and the observation

that unlike text content, video content consumption is not being limited by literacy barriers. On the other hand, the electromagnetic spectrum is a limited resource. In order to ensure that mobile networks continue functioning in a healthy state despite this incredible growth, communication service providers (CSPs) will be required to make infrastructure investments such as more licensed spectrum, cell densification, massive MIMO etc. In order to flatten the rate of growth, CSPs in several markets attempt to identify and shape video traffic based on user data plans. There are several problems with this kind of shaping:

1. Traffic detection and shaping for every flow is compute intensive for CSPs. With distributed UPF (user plane function) in 5G mobile networks more nodes in CSP network may need to support traffic detection and shaping.

2. User mobility during the ongoing session, mainly with distributed UPF (user plane function) in 5G mobile networks may result in shpaing inaccuracies.

3. Traffic detection can have inaccuracies and these inaccuracies are expected to increase as the content delivery industry moves towards end-2-end encryption like TLS 1.3 and encrypted client hello (ECH).

4. The unpredictable behavior of traffic shapers used by CSPs confuse the bandwidth estimation and congestion control protocols being used within end-2-end video delivery sessions between content server and client. This results in poor quality of experience (QoE) for the end user.

5. Content and Application Providers (CAPs) are designing algorithms to detect the presence of such traffic shapers to counter their detrimental effects. These algorithms have their own inaccuracies in detection and add compute resources on the CAP side.

A secure in-band data sharing interface between CSPs and CAPs can enable the CSPs to specify video traffic characteristics (that are suited to their radio access networks) to the CAPs. CAPs can use this information to self-adapt their video traffic to conform to the specified characteristics. Self Adaptation and Self limitation is a better alternative because CAPs have full context and ability to measure user QoE, which CSPs do not. This approach has the potential to help CSPs manage the traffic on their cellular networks without incurring the cost and compute associated traffic detection and traffic shaping while making sure that end user QoE is not compromised which is win-win for both CSPs and CAPs.

What follows are the primary, secondary, and non-requirements of a technical solution to this problem on the modern Internet.

2. Video Optimization Use Case - Primary requirements

2.1. In-band cryptographic key establishment w/ standard crypto

A core requirement is encryption of the data being exchanged between the client endpoint and CSP network device endpoint. In order to achieve this there needs to be a way to have a shared key. This must be done with an in-band mechanism, since out-of-band mechanisms for key exchange are not scalable given the fanout of content providers to CSPs.

2.2. Encryption and integrity protected

A core requirement is the encryption and integrity protection of the data exchanged between the client and CSP network device endpoints. This is crucial to ensure the information cannot be passively observed or modified. Further this encryption must be done with standard ciphers.

2.3. In-band key exchange

In order for encryption to be viable, the client and CSP network device endpoints must have a way to establish a shared key. This mechanism must be done in-band with the network video flow, e.g. via a TLS handshake, so as to avoid the scalability and security problems of sharing keys via an out-of-band mechanism.

2.4. Client-initiated

What is minimally needed is a way for the client to establish a communication channel with a CSP network device, exchange the information, and then use that information to inform its video playback decisions. This also allows for a client device to be aware that the exchange is happening (at least on initiation).

2.5. Low frequency information/data exchange

Video optimization requires a CSP network device to send the allowed data rate for a specific connection to the client endpoint during connection setup time and whenever there is a change in video policy for the subscriber. Change in video policy configuration for a particular subscriber is typically triggered when the subscriber has exhausted its monthly or daily allowed data volume usage limit, i.e. at low frequency.

2.6. Data/Information flows from CSP mobile network to client

There are following two options w.r.t data flow direction to support video optimization use-case. 1. From CSP mobile network to client endpoint 2. From CSP mobile network to CAP server endpoint Both the options have pros and cons. We are proposing option # i due to following reasons; 1. Streaming video flows are predominantly downlink so information can be sent together with downlink packet. There is no need to wait for Uplink QUIC acknowledgements. 2. Communication between CSP mobile network to client endpoint happens over CSP infrastructure which is relatively more secure compared to infrastructure between CSP network device and CAPs' server.

2.7. Scalable for potentially every video flow in a mobile network

This use case requires that potentially every video flow in a mobile network be able to utilize this feature. Thus, it must be performant both for the network device and the mobile device utilizing it.

2.8. Works with QUIC and HTTP/3

HTTP/3 is being used widely as a delivery mechanism for video content by video content providers, and is a critical requirement to support. HTTP/3' s use of QUIC has convenient properties (notably in its use of UDP) that makes solutions in this space more convenient.

2.9. Network device in CSP mobile network: 4G/5G packet core

"Packet core user plane node" is the network device to share information with client endpoint. These nodes are P-GW (PDN Gateway) and UPF(User Plane Function) for 4G and 5G mobile networks respectively. The reasons behind the same are as follows; 1. These nodes have access to subscriber policy via standard 3GPP interface to PCRF (Policy and Charging Rule Function). 2. These nodes are co-located with PCEF (Policy and Charging Enforcement) which is supposed to enforce subscriber specific policy to data flows. 3. Traffic detection function or DPI(Deep Packet Inspection) engine is integrated with these nodes to detect a specific flow/subscriber

2.10. Scalable solution for 4G and 5G mobile packet core from performance standpoint

To support video optimization use-case at scale, significant additional compute in the packet core should not be required. This requirement has some dependency on following aforementioned requirements: 1. As specified earlier in the document, due to low frequency information exchange requirement, there may not be a need for significant additional compute in the packet core to support this

video optimization use-case' s requirements at scale. 2. No need to support traffic shaping in the packet core. This would also free up computational resources.

3. Secondary requirements

1. Works with TCP video flows.

4. Non-requirements

1. Non-HTTP video support.

2. Data flows from CSP mobile network device to CAP server

3. Fixed networks - Fixed network is out of scope at present, since most of the CSPs don' t do video flow shaping for fixed networks. If and when we include fixed networks in the scope, CSP network devices can be CMTS for Cable modem network or BNG for Fiber/DSL network.

5. Information element requirements

This section captures the requirements of information elements to be exchanged between CSP network device and client endpoint of the CAP application. 1. The attributes of video data traffic specified in the information elements - shall be measurable by both CSPs and CAPs. 2. For a given video session the specification - shall ensure that CSP and CAP, albeit measuring independently, compute consistent attributes of the video data traffic. 3. The attributes of video data profile - shall include an average or median video bit rate and a maximum video bitrate 4. The information element - shall specify a methodology for computing median, maximum and average video bitrates including how to determine the time window for measuring these statistics

6. Security Considerations

This document has no security considerations.

7. IANA Considerations

This document has no IANA actions.

Authors' Addresses

Matt Joras
Meta Platforms, Inc.
Email: matt.joras@gmail.com

Anoop Tomar
Meta Platforms, Inc.
Email: anooptomar@meta.com

Abhishek Tiwari
Meta Platforms, Inc.
Email: atiwari@meta.com

Alan Frindell
Meta Platforms, Inc.
Email: afrind@meta.com