Mobile Throughput Guidance
Version 1.0
08 March 2017

This is a White Paper of the GSMA

Security Classification: Non-confidential

Copyright Notice

Disclaimer

Antitrust Notice

Table of Contents

1 Introduction

1.1 Overview

This document presents Mobile Throughput Guidance (MTG) as a potential means to improve customer experience during mobile Internet sessions, by making explicit what range of bandwidth the mobile access link is likely to sustain. The document recommends that GSMA operators and vendors support the activity through further investigation, with the goal of contribution towards an IETF standard.

1.2 Abbreviations

Term	Description
3G	3rd Generation Mobile Network
ARQ	Automatic Repeat Request (L1)
GPU	Graphical Processing Unit
HARQ	Hybrid Automatic Repeat Request (L2)
HSPA	High Speed Packet Access
ICCRG	Internet Congestion Control Research Group
IETF	Internet Enginnering Task Force
LTE	Long Term Evolution
MTG	Mobile Throughout Guidance. A network-calculated information element which recommends a sustainable bandwidth to flow endpoints.
PLUS	Path Layer UDP Substrate
TCP	Transmission Control Protocol
TLS	Tarnsport Layer Security
UDP	User Datagram Protocol

1.3 References

Ref	Doc Number	Title
[1]	CC-4G-5G	Ingemar Johansson. Available at https://tools.ietf.org/html/draft-johansson-cc-for-4g-5g-02
[2]	ACCORD	“An Internet perspective of 3GPP architecture”, K. Smith, IETF96 ACCORD BoF, April 2016. Available at https://www.ietf.org/proceedings/95/slides/slides-95-accord-2.pdf
[3]	MTG-PCL	“Mobile Throughput Guidance Inband Signaling Protocol”, H. Flinck et al., September 2015. Available at https://tools.ietf.org/html/draft-flinck-mobile-throughput-guidance-03
[4]	MTG-REQ	“Requirements and reference architecture for Mobile Throughput Guidance”, N. Sprecher et al., March 2016. Available at https://tools.ietf.org/html/draft-sprecher-mobile-tg-exposure-req-arch-02
[5]	RFC 2119	“Key words for use in RFCs to Indicate Requirement Levels”, S. Bradner, March 1997. Available at http://www.ietf.org/rfc/rfc2119.txt

2 Problem statement

2.1 The volatile nature of cellular radio access

Mobile throughput can vary considerably for a given mobile connection during a data session. Significant factors that cause this volatile throughput include:

- The radio bearer. LTE data capacity and throughput exceeds HSPA and 3G, for example.
- Received signal quality as continuously reported by the client device baseband, which indicate:
    - The signal to noise ratio. Environmental conditions and physical objects cause interference, which can change as the client moves around an area.
    - The signal strength will fade as the distance to the source increases
- Mobility between radio access nodes results in handover of buffers between the source and target nodes, which can result in loss and delay. 3G systems apply ‘soft handover’, in which the flow is replicated in parallel on the target node, and then switched off on the source node once attachment to the target node confirmed. LTE systems apply ‘hard handover’, where the source node flow is switched off and then immediately routed to the target node. Soft handover is safer with the 3G capacity constraints, because of the redundancy of the second flow, but it does mean duplication of flows and hence adds to the source scheduler load. LTE capacity means hard handover is viable, but can result in loss if the source buffers are large and need to be handed off to the target node.
- The network buffer configuration. Operators may configure buffers with high capacity to ensure that radio schedulers always have data to allocate to radio blocks, thus ensuring efficient use of spectrum. This may result in buffer bloat and hence packet drops or delay at the network queues.
- Congestion at the radio scheduler. This places incoming content into available radio blocks, determined by device and load. High load on the scheduler leads to resource contention and queuing or loss, or colloquially, “a congested cell”.

The radio throughput is a compound of the above conditions: in short, the quality of the radio connection, the load on the radio access network, and the handover of flows and queues between access nodes. Please see [Ref: 2] for further details.

The nature of the data flow is also a factor. A frequent, low bitrate flow is easier for the radio channel to maintain at a constant rate than a flow with periodic bursts, such as Adaptive Bitrate Video. As device screen resolutions and device GPU capabilities improve, this drives content providers to produce higher quality streams, which in turn places a greater load on the network as chunks of high quality video are requested in bursts.

2.1.1 What problems does this cause?

Customer experience can be degraded as a direct result of volatile radio throughput. This is more likely to manifest in flows requiring high capacity and low latency, because such conditions cannot be guaranteed for the lifecycle of the session.

TCP-like congestion control algorithms are typically loss-based: if the TCP endpoints at the client and server infer packet loss, the server will drastically reduce the sending rate because it assumes congestion is the cause. However, the loss inference is based on acknowledgement timers: i.e. if a TCP segment is not acknowledged as received within a certain time. Although LTE has robust loss recovery at Layer 1 (ARQ) and rapid loss recovery at Layer 2 (HARQ), in poor radio conditions in can take several radio retransmissions, and hence delay, to deliver the TCP segment to the client. Therefore this delay can exceed the TCP acknowledgement timer, and result in an incorrect perception of congestion.

When recovering from perceived congestion, and the ‘multiplicative decrease’ of sending rate, TCP will carefully probe the network with a far smaller sending rate which gently increases (‘additive increase’). This means that where a device is moving to better quality radio conditions, which can happen quickly especially in urban areas due to reflections and absence of physical barriers, then the now-available strong connection will not be utilised quickly: rather it will take several round trips to ‘get up to speed’.

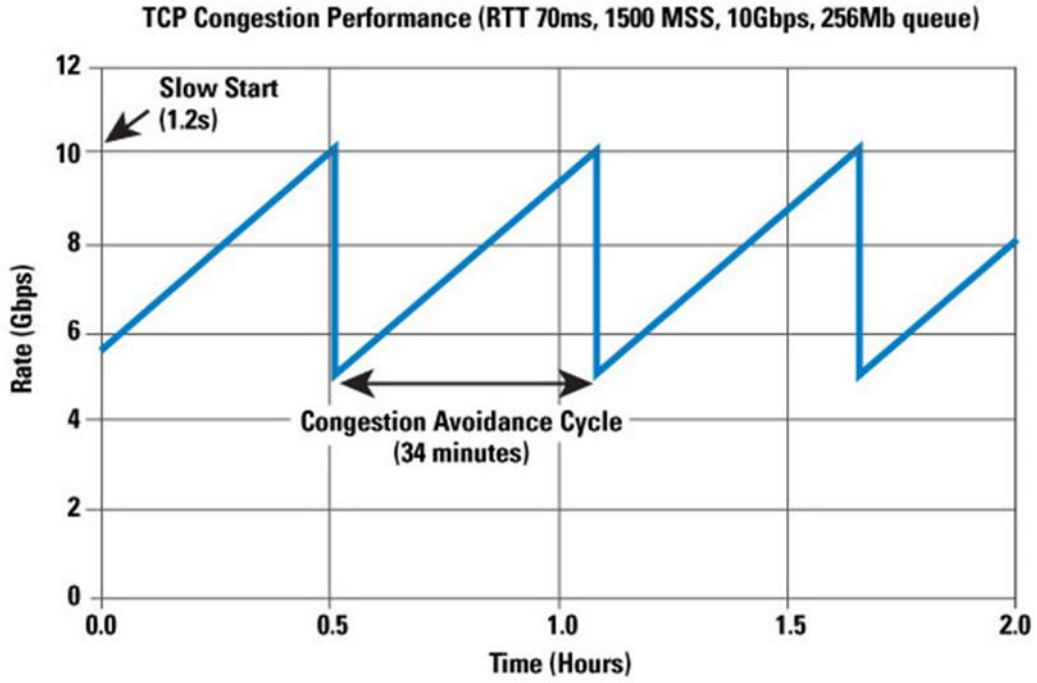

Figure 1: TCP at high speed, “Gigabit TCP’, the Internet Protocol Journal vol. 9 no. 2, G. Houston, 2006

Overall the resulting variance in radio conditions, and the effect on endpoint congestion controls, can result in jitter (variable rates of packet pacing), stalling (as buffers fill too quickly) and hence a poor customer experience, especially for real-time or video streaming services. Please see [Ref:1] for further details.

3 Solution design

For details on solution architecture principles, requirements and use cases, please see [Ref: 4]. These are summarised below.

3.1 Requirements

- MTG may be provided per flow or per user. The latter implies that the MTG take into account multiple flows for that user, because if the user has multiple high-bandwidth flows there will be an impact on device buffers as well as the network.
- Encrypted traffic (e.g. TLS over TCP) is supported.
- The injection of MTG should not degrade the TCP flow performance.
- Middleboxes should not amend/remove the MTG.
- The consumer of the information may choose to act on the guidance.

3.2 Use case: mobile video delivery optimisation

A video utilising MTG will have more contextual information with which to set its initial congestion window and make in-flow adjustments. This can therefore remove the need for slow start and buffering, and can allow more consistent pacing throughout the video playback.

3.2.1 Network Security considerations

Integrity and authentication requirements for the protocol [Ref:4]. Network operators will also need to ensure that the exposure of MTG does not leak business sensitive or network security information. This includes any information which can identify the network entities involved in delivering the guidance; and information that can reveal the performance of a radio access node over a given period. Constraints on the information model and the frequency of guidance injection are therefore recommended.

3.2.2 Customer privacy considerations

MTG indicates the range of throughput that can be supported for a TCP connection. No customer identifiers are included in the information set. Should MTG be compromised, the effect is likely that (1) it be considered useless by the endpoints, and (2) that it would only reveal an approximate state of a part of the operator network, likely to be identical for other clients utilising that network route.

4 Mobile throughput guidance

This section summarises [Ref:4] and a proposed evolution of that draft specification, as discussed between contributors.

4.1 Information model

- The version identifier of the guidance specification adhered to.
- The approximate congestion level of the radio access, as a result of analysing network buffer state.
- The suggested throughput that the endpoints should aim for.

4.2 Protocol model

[Ref:4] details the TCP binding, including the option to negotiate an encrypted mode.

4.3 Constraints

- The solution is currently bound to TCP, meaning no support for other transport protocols.
- Guidance implies a safe range rather than hard accurate values.
- Guidance can at best reflect radio access conditions up to the point the guidance is injected. The operator network beyond this point, through the core network and up to the Internet interface and firewalls, is not considered in evaluating the guidance.
- Throughput is volatile enough to change within the time taken to write the value and forward to the consumer of the information (one or both of the flow endpoints). The value of the guidance therefore diminishes as latency of its delivery increases.
- At IETF 93, the MTG authors presented to the ICCRG (Internet Congestion Control Research Group). There were concerns raised that the result of MTG may be to simply move bottlenecks, rather than resolve them (see 'Recommendations' below).

4.4 Test results

The video optimisation results available at [MTG-REQ] involve a live LTE network. These show a significant reduction (~20%) in re-buffering time and an increase (>5%) in video resolution.

5 Recommendation

5.1 Standards contribution

Although the IETF internet drafts for [Ref:4] and [Ref:3] have expired, discussions among the contributors are still active with an intention to present a revised version at an upcoming (2017) IETF meeting. As well as the revised protocol model, the following areas will need to be addressed:

- Support for other protocols, since the current binding is for TCP only. Note that MTG at the IP layer is considered problematic, as there is no clear area of the IPv4 or IPv6 header to place the information without (1) potential conflict or misinterpretation by other processes or (2) dropping the information by interim routers that are not configured to expect it. The recommendation is that if PLUS (Path Layer UDP Substrate) becomes a viable means to transfer information from the network to endpoints via a UDP 'signalling plane', then MTG can be transferred over PLUS.
- Ensure no shift of bottleneck. This will require further discussions with ICCRG

5.2 Operator contribution

Operators are encouraged to engage with the MTG activity through review of internet drafts and live trials, sharing (anonymised and approved) results where possible.

5.3 Implementation considerations

Operator network security and privacy teams should review implementation plans. As with any information in a privacy context, MTG may seem non-customer-sensitive in IETF, but could be combined with other information about the flow to build a fuller picture of customer behaviours.

Annex A Document Management

A.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	7/12/16	Document for SMART sub-group review	IG	Kevin Smith, Vodafone

A.2 Other Information

Type	Description
Document Owner	Internet Group
Editor / Company	Kevin Smith, Vodafone

It is our intention to provide a quality product for your use. If you find any errors or omissions, please contact us with your comments. You may notify us at prd@gsma.com

Your comments or suggestions & questions are always welcome.