Requirements and reference architecture for Mobile Throughput Guidance
                               Exposure
             draft-sprecher-mobile-tg-exposure-req-arch-03.txt

Abstract

   Rapidly-varying conditions in a cellular network can cause problems
   for the Transmission Control Protocol (TCP), which in turn can
   degrade application performance.

   This document presents the problem statement and proposes solution
   principles. It specifies the requirements and reference architecture
   for a mobile throughput guidance exposure mechanism that can be used
   to assist TCP in cellular networks, ensuring better network
   efficiency and enhanced service delivery performance.

   The proposed mechanism can be applied to any content or TCP/IP based
   application content delivery. This document describes the
   applicability of the mechanism for Intelligent Video Acceleration
   over cellular networks.

Status of This Memo

   This Internet-Draft will expire on September 14, 2017.

Copyright Notice

   Copyright (c) 2017 IETF Trust and the persons identified as the
   document authors. All rights reserved.

   This document is subject to BCP 78 and the IETF Trust's Legal
   Provisions Relating to IETF Documents
   (http://trustee.ietf.org/license-info) in effect on the date of
   publication of this document. Please review these documents
   carefully, as they describe your rights and restrictions with respect
   to this document. Code Components extracted from this document must
   include Simplified BSD License text as described in Section 4.e of
   the Trust Legal Provisions and are provided without warranty as
   described in the Simplified BSD License.

Table of Contents

1. Introduction

   Mobile-edge Computing provides IT and cloud-computing capabilities
   within the Radio Access Network in close proximity to mobile
   subscribers [MEC_White_Paper]. Mobile Throughput Guidance is an
   solution building on top of MEC concepts to improve customer
   experience during mobile Internet sessions, by making explicit what

range of bandwidth the mobile access link is likely to sustain
[MTG_GSMA_White_paper].

The following sub-sections present the problem statement and the
solution principles for Mobile Throughput Guidance.

1.1. Contributing Authors

The editors gratefully acknowledge the following additional
contributors: Hannu Flinck, Helen Parsons, Peter Cosimini and Ram
Gopal.

1.2. 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 RFC 2119 [RFC2119].

1.3. Acronyms and Abbreviations

HTTP Hypertext Transmission Protocol
IP   Internet Protocol
LTE Long Term Evolution
RAN Radio Access Network
RTT Round Trip Time
TCP Transmission Control Protocol
UE   User Equipment

1.4. Problem statement

Inefficient use of a cellular network's resources degrades
application performance, delivery of content and user experience.

Cellular networks are often required to deliver large, high bandwidth
files to end users, e.g from streaming media content providers. If
the available throughput from the Radio Access Network (RAN) to the
User Equipment (UE) falls below the bandwidth required then files are
delivered too slowly, resulting in a bad user experience. It may be
possible to take avoiding action and so limit the impact on the
network and the user experience. However to be able to do this in an
accurate and timely fashion, information on the available throughput
is required.

Internet media and file delivery are typically streamed or downloaded
today using Hypertext Transmission Protocol (HTTP) over the TCP. The
behavior of TCP assumes that network congestion is the primary cause
for packet loss and high delay. This may not be the case in cellular
networks where the bandwidth available for each UE can vary by an

order of magnitude within a few seconds due to changes in the
underlying radio channel conditions. Such changes can be caused by
the movement of devices or interference, as well as changes in system
load due to bursty traffic sources or when other devices enter and
leave the network. On the other hand, packet losses tend to be
sporadic and temporary; retransmission mechanisms at the physical and
link layers repair most packet corruptions.

1.5. Solution Principles

   This document proposes that the cellular network could provide near
   real-time information on "Throughput Guidance" to the TCP server;
   this throughput guidance information would indicate the throughput
   estimated to be available at the radio downlink interface (between
   the RAN and the UE) for the TCP connection.

   While the implementation details will vary according to the cellular
   access network technology, the resource allocation can be abstracted
   as the capacity of the "radio link" between the network and the UE.
   For example, in the case of an LTE network, the number of physical
   resource blocks allocated to a UE, along with the modulation scheme
   and coding rate used, can be translated into radio link capacity in
   Megabits per second (Mbps). It can also include the quality of the
   "radio link" which is reported by the UE.

   The TCP server can use this explicit information to inform several
   congestion control decisions. For example: (1) selecting the initial
   window size, (2) deciding the value of the congestion window during
   the congestion avoidance phase, and (3) reducing the size of the
   congestion window when the conditions on the "radio link"
   deteriorate. In other words, with this additional information, TCP
   does neither have to congest the network when probing for available
   resource, nor rely on heuristics to reduce its sending rate after a
   congestion episode.

   The same explicit information can also be used to optimize
   application behavior given the available resources. For example,
   when video is encoded in multiple bitrates, the application server
   can select the appropriate encoding for the network conditions.

   Note that the throughput estimation for the upstream traffic between
   the UE and the RAN, and the throughput of the network path between
   the RAN and the server communicating with the UE are beyond the scope
   of the document.

   It is also important to note that the validity of the throughput
   guidance and the distance between the originating server and the
   cellular network (in terms of the number of Internet hops) are

inversely proportional. This is due to the fact that the latency
incurred at each hop increases the time that elapses between issuing
and consuming the guidance.

2. Requirements

The requirements set out in section 2.1 are for the behavior of the
mobile throughput guidance exposure mechanism and the related
functional elements. The related security requirements are specified
in section 2.2.

2.1. Requirements on the Mobile Throughput Guidance Exposure Mechanism

   1.   The throughput guidance information SHALL indicate the expected
        available bandwidth in the downlink interface. Depending on the
        solution mechanism, the information MAY be provided per TCP flow
        or per user. If the solution mechanism supports both options,
        then granularity SHOULD be configurable.

   2.   The throughput guidance information SHALL be provided for TCP
        based traffic.

   3.   A functional element, residing in the RAN and acting as
        Throughput Guidance Provider, SHOULD supply the TCP server a
        near real-time indication (in sub-seconds) on the throughput
        estimated to be available at the radio downlink interface (i.e.,
        mobile throughput guidance information). It SHOULD keep up with
        the rapid changes in the radio network conditions, the network
        traffic and the user movement, in order to provide the most
        accurate guidance information. Throughput Guidance Provider
        SHALL access to Network Radio Information.

   4.   The introduction of the Throughput Guidance exposure mechanism
        SHALL NOT require any update to the TCP client software.

   5.   The mobile throughput guidance exposure mechanism SHALL work
        when the user traffic is end-to-end encrypted (e.g., HTTPS,
        etc.). This requirement is compliant with the IAB Statement on
        Internet Confidentiality (see [IAB_Statement]), saying that the
        IAB "strongly encourage developers to include encryption in
        their implementations, and to make them encrypted by default.
        We similarly encourage network and service operators to deploy
        encryption where it is not yet deployed, and we urge firewall
        policy administrators to permit encrypted traffic.".

   6.   The mobile throughput guidance exposure mechanism SHALL NOT
        adversely impact the behavior of the TCP flows (e.g., it SHOULD

NOT cause an increase in retransmissions or degradation in performance, etc.).

7. The mobile throughput guidance exposure mechanism SHALL ensuring end-to-end consistency of the TCP flow control. The throughput guidance information SHALL NOT override the TCP flow window limits (e.g. min(cwnd, rwnd)) or timer values.

8. The throughput guidance information SHALL be opaque to the intermediate elements between the Throughput Guidance Provider and the TCP server. The intermediate elements SHOULD NOT modify or remove the throughput guidance information.

9. The TCP server MAY reside within the mobile operator's network (behind the mobile core network) or in the Internet.

10. The TCP server MAY use the mobile throughput guidance information to assist TCP.

11. It SHOULD be possible for the TCP server to provide the exposed mobile throughput guidance information to an authorized higher layer application. The application may use the mobile throughput guidance information to optimize its behavior.

12. The Throughput Guidance Provider SHOULD provide the mobile throughput guidance information periodically, starting from the initiation of the flow.

13. The frequency (in milliseconds) at which mobile throughput guidance needs to be exposed SHALL be configurable.

14. The mobile Throughput Guidance Provider SHALL be able to supply mobile throughput guidance information to more than one TCP server simultaneously, with independent configurable parameters for each server.

15. There SHOULD be a mechanism to configure the Mobile Throughput Guidance Provider with a list of TCP flows for which mobile throughput guidance information shall be exposed.

16. The mobile throughput guidance exposure mechanism SHOULD ensure backward compatibility. Normal TCP processing at the TCP server SHOULD be performed if the TCP server does not recognize the throughput guidance information.

17. The mobile throughput guidance exposure mechanism MUST be extensible, ensuring that additional information can be provided in the future in a non-disruptive, backward-compatible way.

2.2. Security requirements

   1. A trustful relationship between the Mobile Throughput Provider
       and the TCP server SHOULD be formed before any information is
       exposed.

   2. There SHOULD be a mechanism to configure the Mobile Throughput
       Guidance Provider with a list of destinations to which throughput
       guidance should be provided.

   3. The identity of the Mobile Throughput Guidance Provider SHALL be
       explicitly known to the TCP server which receives the
       information. The TCP server SHALL be able to authenticate the
       identity of the Mobile Throughput Guidance Provider. The Mobile
       Throughput Guidance Provider MUST NOT reveal any other identity
       or address of network elements that can compromise the security
       of the network.

   4. The mobile throughput guidance information SHOULD be secured to
       ensure confidentiality and integrity.

   5. There SHOULD be a mechanism to configure the required security
       level and parameters for the encryption and the authentication if
       supported.

   6. The exposure of the Mobile throughput guidance information SHALL
       NOT introduce any additional security threats and privacy
       concerns to the mobile operator's network, the Internet and the
       users.

   7. The throughput guidance SHOULD be treated only as an estimate to
       the optimization algorithm running at the TCP server. The TCP
       server that receives this information SHOULD NOT assume that it
       is always accurate and up to date. Specifically, the TCP server
       SHOULD check the validity of the information received and if it
       finds it erroneous it SHOULD discard it and possibly take other
       corrective actions (e.g., discard all future throughput guidance
       information from a particular IP prefix).
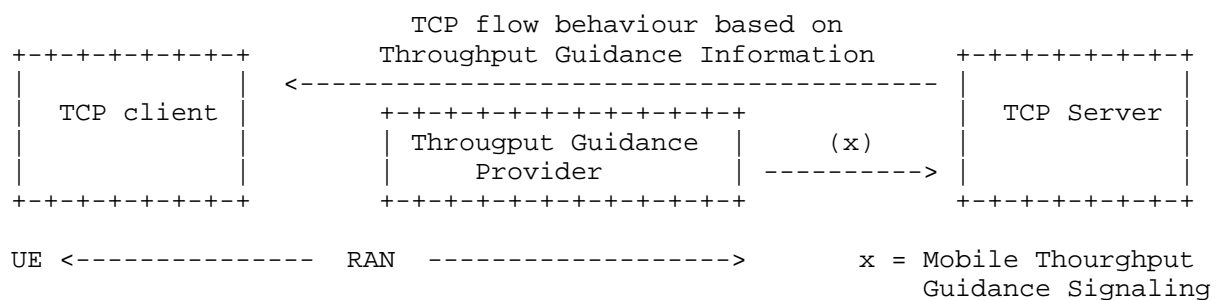
3. Reference Architecture

   Figure 1 below, depicts the functional elements and their interfaces
   that comprise the mobile network guidance solution (based on the
   requirements for mobile throughput guidance).

   A Throughput Guidance Provider functional element signals to the TCP
   server the information on the (near-real time) throughput estimated

to be available at the radio downlink interface. The TCP server
resides within the mobile operator's network or in the Internet.

Note that the Throughput Guidance Provider functional element and the
TCP server can belong to the same Administrative System (AS) or to
different Administrative Systems.

The TCP server MAY use the information to optimize the TCP behavior.
The information MAY also be used by the application to adapt its
behavior accordingly and to optimize service delivery performance.


```
                          TCP flow behaviour based on
+-+-+-+-+-+-+-+          Throughput Guidance Information       +-+-+-+-+-+-+-+
|           |   <--------------------------------------- |           |
|  TCP client |         +-+-+-+-+-+-+-+-+-+-+-+         | TCP Server |
|           |         | Througput Guidance |    (x)    |           |
|           |         |      Provider      | ----------> |           |
+-+-+-+-+-+-+-+         +-+-+-+-+-+-+-+-+-+-+-+         +-+-+-+-+-+-+-+

UE <-------------- RAN ------------------->        x = Mobile Thourghput
                                                       Guidance Signaling
```


              Mobile Throughput Guidance Reference Architecture

                               Figure 1

The information source and the algorithm used by the Throughput
Guidance Provider to calculate the throughput guidance are beyond the
scope of this document.

The TCP server MAY use the throughput guidance information to assist
TCP in any of the following ways:

o Determine the size of the initial congestion window

o Determine when to exit the slow start phase

o Determine the size of the congestion window during the congestion
   avoidance phase

o Determine the size of the window after a congestion event

4. Applicability to Mobile Video Delivery Optimization

   The mobile throughput guidance exposure mechanism applies to mobile
   video delivery optimization.

   In this use case the Throughput Guidance Provider sends to the video
   server throughput guidance information for a TCP flow. The video
   server may use this information to assist TCP congestion control
   decisions, for example in selecting the initial congestion window
   size, and adjusting the size of the congestion window when the
   conditions on the radio link change. In other words, with this
   additional information, TCP does not need to overload the network
   when probing for available resources, nor does it need to rely on
   heuristics to reduce its sending rate after a congestion episode.
   Slow start and buffering of content delivery can be eliminated.

   The same information may also be used to ensure that the application
   level coding matches the estimated capacity at the radio downlink.

   The aim of all of these improvements is to enhance the end user's
   quality of experience. For example, the content's time-to-start as
   well as video buffering occurrences can be reduced, the utilization
   of the radio network's resources and its throughput can be optimized,
   etc.

5. Manageability considerations

   Manageability of mobile throughput guidance exposure will be
   discussed in the solution documents. Section 2 specifies a set of
   requirements on the management of the mobile throughput guidance
   exposure functional elements and protocol operation.

6. Security considerations

   The exposure of mobile throughput guidance information from the
   cellular network to the TCP server introduces a set of security
   considerations.

   As per requirement #3 in section 2.2, the TCP server SHALL be able to
   authenticate the identity of the Mobile Throughput Guidance Provider.
   The Mobile Throughput Guidance Provider MUST NOT reveal any other
   identity or address of network elements that can compromise the
   security of the network.

   Furthermore, the throughput guidance information should be treated
   only as an estimate to the congestion control algorithm running at
   the transport endpoint. The endpoint that receives this information
   should not assume that it is always correct and accurate.

Specifically, endpoints should check the authenticity and integrity of the information received and if they find it erroneous they should discard it and possibly take other corrective actions (e.g., discard all future throughput guidance information from a particular IP prefix).

One way to check if the throughput guidance information overestimates the capacity available on the radio link is to check whether any packet losses or other signs of congestion (e.g., increasing RTT) occur after the guidance is used. Notably, the same mechanism can be used to deal with bottlenecks in other parts of the end-to-end network path. To check if the throughput guidance underestimates the available network capacity, the source can periodically attempt to send faster and then check for signs of congestion.

Section 2 above, specifies a set of requirements on the mobile throughput guidance exposure protocol to ensure secured communication and operation.

7. IANA considerations

This requirements and architecture document does not introduce any requests for IANA actions.

8. Acknowledgements

We would like to thank Peter Szilagyi, Meir Cohen and Csaba Vulkan for conversations on these issues.

9. References

9.1. Normative References

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7,
           RFC 793, DOI 10.17487/RFC0793, September 1981,
           <http://www.rfc-editor.org/info/rfc793>.

[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>.

[RFC6994] Touch, J., "Shared Use of Experimental TCP Options",
           RFC 6994, DOI 10.17487/RFC6994, August 2013,
           <http://www.rfc-editor.org/info/rfc6994>.

9.2. Informative References

   [I-D.narten-iana-considerations-rfc2434bis]
             Narten, T. and H. Alvestrand, "Guidelines for Writing an
             IANA Considerations Section in RFCs", draft-narten-iana-
             considerations-rfc2434bis-09 (work in progress), March
             2008.

   [IAB_Statement]
             "IAB Statement on Internet Confidentiality", November
             2014, <http://www.ietf.org/mail-archive/web/ietf-
             announce/current/msg13460.html>.

   [MEC_White_Paper]
             ETSI, "Mobile-Edge Computing - Introductory Technical
             White Paper", September 2014,
             <http://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-
             edge_Computing_-
             _Introductory_Technical_White_Paper_V1%2018-09-14.pdf>.

   [MTG_GSMA_White_paper]
             GSMA, "Mobile Throughput Guidance White Paper", March
             2017, <https://portal.etsi.org/portals/0/tbpages/mec/docs/
             mobile-edge_computing_-
             _introductory_technical_white_paper_v1%2018-09-14.pdf>.

   [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629,
             DOI 10.17487/RFC2629, June 1999,
             <http://www.rfc-editor.org/info/rfc2629>.

   [RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC
             Text on Security Considerations", BCP 72, RFC 3552,
             DOI 10.17487/RFC3552, July 2003,
             <http://www.rfc-editor.org/info/rfc3552>.

   [RFC4413] West, M. and S. McCann, "TCP/IP Field Behavior", RFC 4413,
             DOI 10.17487/RFC4413, March 2006,
             <http://www.rfc-editor.org/info/rfc4413>.

Authors' Addresses

Ankur Jain
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Phone: +1-925-526-5879
Email: jankur@google.com

Andreas Terzis
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Phone: +1-650-214-5270
Email: aterzis@google.com

Nurit Sprecher
Nokia Networks
Hod HaSharon
IL

Phone: +97297751229
Email: nurit.sprecher@nsn.com

Swaminathan Arunachalam
Nokia Networks
Irving
FUS

Phone: +19723303204
Email: swaminathan.arunachalam@nsn.com

Kevin Smith
Vodafone
One Kingdom Street, Paddington Central
London W2 6BY
UK

Email: kevin.smith@vodafone.com

Guenter Klas
Vodafone
One Kingdom Street, Paddington Central
Newbury RG14 2FN
UK

Email: kguenter.klas@vodafone.com