ETSI GR NGP 010 V1.1.1 (2018-09)

Next Generation Protocols (NGP); Recommendation for New Transport Technologies

Disclaimer

The present document has been produced and approved by the Next Generation Protocols (NGP) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference

DGR/NGP-0010

Keywords

Next Generation Protocol, QoS, transport

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Next Generation Protocols (NGP).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

The present document focuses on new transport technology for next generation architectures toward 5G and beyond. The basic concept is to enhance the best-effort based IP network to QoS capable IP network. The goal is to provide the QoS for the upper layer protocols. The work aims to examine and propose recommendations to improve and simplify the network infrastructure to support QoS for different transport protocols. In addition, the present document may require the development of new protocols and or modification of existing protocols.

Introduction

Recently, more and more new applications for Internet are emerging. These applications have a common requirement to the Internet that is their required bandwidth is very high and/or latency is very low compared to traditional applications like most of web browser and video streaming applications.

For example, AR or VR applications may need at least couple of hundred Mbps bandwidth (throughput) and a low single digit MS latency. Moreover, the difference of mean bit rate and peak bit rate is huge due to the compression algorithm [i.1].

Some future applications expect that Internet can provide a up bounded latency (minimized latency) service, such as tactile network [i.2]. To these applications, the latency will determine their user experience or application quality, so it is critical that the maximum latency for application is bounded within values application has requested.

With the technology development in 5G and beyond, the wireless access network is also rising the demand for the Ultra-Reliable and Low-Latency Communications (URLLC), this also leads to the question if IP transport can provide such service in Evolved Packet Core (EPC) network. IP is becoming more and more important in EPC when the Multi-access Edge Computing (MEC) for 5G will require the cloud and data service moving closer to eNodeB.

The present document will brief the current IP transport and QoS technologies, and analyse the limitations to support above new applications.

A frame work for new transport technology based on QoS enabled IP network will be reported. As an example, detailed design and experiments for TCP are given.

The frame work also lists other areas, topics and issues that need more study to achieve the complete solution.

1 Scope

The present document reports the analysis of current transport technologies for Internet, especially TCP, the limit of different variants for TCP and other transport protocols, and then proposes a framework for new transport technology for IP network. TCP is exemplified for the detailed design and prove of concept experiments.

In the design, both control plane and data plane are discussed. It includes the control mechanism, message type, key message parameters, hardware capability, forwarding state, host congestion control and traffic management.

In the experiments, the POC product and its realization are discussed; test results, scalability and performance are analysed.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] Draft-han-iccrg-arvr-transport-problem-01 (work in progress): "Problem Statement: Transport Support for Augmented and Virtual Reality Applications", L. Han, and K. Smith, March 2017.

[i.2] Proceedings of European Wireless 2015; 21th European Wireless Conference: "Towards the Tactile Internet: Decreasing Communication Latency with Network Coding and Software Defined Networking", J David Szabo, 2015.

NOTE: Available at https://ieeexplore.ieee.org/iel7/7147658/7147659/07147730.pdf.

[i.3] DEC Research Report TR-301: "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems", R. Jain, 1984.

NOTE: Available at http://www1.cse.wustl.edu/~jain/papers/ftp/fairness.pdf.

[i.4] Andreas Benthin, Stefan Mischke, University of Paderborn: "Bandwidth Allocation of TCP", 2004.

[i.5] IETF RFC 2581: "TCP Congestion Control", M. Allman, V. Paxson and W. Stevens, April 1999.

NOTE: Available at https://www.rfc-editor.org/info/rfc2581.

[i.6] L. Peterson: "TCP Vegas: New Techniques for Congestion Detection and Avoidance - CiteSeer page on the 1994 SIGCOMM paper", 1994.

[i.7] S. Ha, I. Rhee and L. Xu: "CUBIC: A New TCP-Friendly High-Speed TCP Variant", 2008.

[i.8] Draft-sridharan-tcpm-ctcp-02 (work in progress): "Compound TCP: A New TCP Congestion Control for High-Speed and Long Distance Networks", M. Sridharan, K. Tan, D. Bansal and D. Thaler, November 2008.

[i.9] Radhika Mittal, Vinh The Lam, Nandita Dukkipati, Emily Blem, Hassan Wassel, Monia Ghobadi, Amin Vahdat, Yaogong Wang, David Wetherall, David Zats: "TIMELY: RTT-based Congestion Control for the Datacenter", 2010.

NOTE: Available at http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p537.pdf.

[i.10] Draft-falk-xcp-spec-03 (work in progress): "Specification for the Explicit Control Protocol (XCP)", A. Falk, Jul 2007.

[i.11] Nandita Dukkipati, Ph.D. Thesis, Department of Electrical Engineering, Stanford University: "Rate Control Protocol (RCP): Congestion control to make flows complete quickly", 2007.

NOTE: Available at http://yuba.stanford.edu/~nanditad/thesis-NanditaD.pdf.

[i.12] Draft-ietf-tcpm-dctcp-03 (work in progress): "Datacenter TCP (DCTCP): TCP Congestion Control for Datacenters", S. Bensley, L. Eggert, D. Thaler, P. Balasubramanian, and G. Judd, November 2016.

[i.13] Draft-ietf-aqm-pie-10 (work in progress): "PIE: A Lightweight Control Scheme To Address the Bufferbloat Problem", R. Pan, P. Natarajan, F. Baker, and G. White, September 2016.

[i.14] Draft-ietf- aqm-codel-06 (work in progress): "Controlled Delay Active Queue Management", K. Nichols, V. Jacobson, A. McGregor, and J. Iyengar, December 2016.

[i.15] Draft-ietf-aqm-fq-codel-06 (work in progress): "The FlowQueue-CoDel Packet Scheduler and Active Queue Management Algorithm", T. Hoeiland-Joergensen, P. McKenney dave.taht@gmail.com, J. Gettys and E. Dumazet, March 2016.

[i.16] Lavanya Jose, Mohammad Alizadeh, George Varghese, Nick McKeown, Sachin Kattie: "High Speed Networks Need Proactive Congestion Control", 2016.

NOTE: Available at http://web.stanford.edu/~lavanyaj/papers/perc-hotnets15.pdf.

[i.17] Neal Cardwell, Yuchung Cheng, C. Stephen Gunn, Soheil Hassas Yeganeh,Van Jacobson: "BBR Congestion Control", 2016.

NOTE: Available at https://www.ietf.org/proceedings/97/slides/slides-97-iccrg-bbr-congestion-control-02.pdf.

[i.18] Mo Dong, University of Illinois at Urbana-Champaign, Hebrew University of Jerusalem: "PCC: Re-architecting Congestion Control for Consistent High Performance", 2014.

NOTE: Available at https://arxiv.org/abs/1409.7092.

[i.19] Jonathan Perry: "Fastpass: A Centralized "Zero-Queue" Datacenter Network", 2014.

NOTE: Available at http://fastpass.mit.edu/Fastpass-SIGCOMM14-Perry.pdf.

[i.20] Matthew Mathis, Pittsburgh Supercomputing Center: "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", 1997.

NOTE: Available at https://cseweb.ucsd.edu/classes/wi01/cse222/papers/mathis-tcpmodel-ccr97.pdf.

[i.21] Wei Bao, The University of British Columbia, Vancouver, Canada, IEEE Globecom 2010 proceedings: "A Model for Steady State Throughput of TCP CUBIC", 2010.

NOTE: Available at https://www.researchgate.net/publication/224211021_A_Model_for_Steady_State_Throughput_of_TCP_CUBIC.

[i.22] IETF RFC 2475: "An Architecture for Differentiated Services".

NOTE: Available at https://www.rfc-editor.org/info/rfc2475.

[i.23] IETF RFC 1633: "Integrated Services in the Internet Architecture: an Overview".

NOTE: Available at https://www.rfc-editor.org/info/rfc1633.

[i.24] IETF RFC 8200: "Internet Protocol, Version 6 (IPv6) Specification".

NOTE: Available at https://www.rfc-editor.org/info/rfc8200.

[i.25] Draft-harper-inband-signalling-requirements-00 (work in progress): "Requirements for In-Band QoS Signalling", J. Harper, January 2007.

[i.26] Draft-roberts-inband-qos-ipv6-00 (work in progress): "In-Band QoS Signaling for IPv6", L. Roberts and J. Harford, July 2005.

[i.27] IETF RFC 4782: "Quick-Start for TCP and IP".

NOTE: Available at https://www.rfc-editor.org/info/rfc4782.

[i.28] IETF RFC 5971: "GIST: General Internet Signalling Transport".

NOTE: Available at https://www.rfc-editor.org/info/rfc5971.

[i.29] IETF RFC 2460: "Internet Protocol, Version 6 (IPv6) Specification".

NOTE: Available at https://www.rfc-editor.org/info/rfc2460.

[i.30] IETF RFC 6275: "Mobility Support in IPv6".

NOTE: Available at https://www.rfc-editor.org/info/rfc6275.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

deterministic IP: term contrast to best-effort IP and intend to represent new IP that has QoS support for bandwidth and minimum latency

NOTE: It is similar to the objectives of IETF Detnet WG.

in-band signaling: control information sent within the same band or channel used for user data

IP flow: data flow identified by the source, destination IP address, the protocol number, the source and destination port number

IP path: route that IP flow will traverse

NOTE: IP path could be the shortest path determined by routing protocols (IGP or BPG), or the explicit path decided by another management entity, such as a central controller, or Path Computation Element (PCE) Communication Protocol (PCEP), etc.

out-of-band signaling: control information sent over a different channel, or even over a separate network

QoS channel: forwarding channel that the QoS is guaranteed so to provide additional QoS service to the normal IP forwarding

NOTE: A QoS channel can be used for one or multiple IP flows depends on the granularity of in-band signaling.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACK	Acknowledge
ACL	Access Control List
AIMD	Additive-Increase/Multiplicative-Decrease

API	Application Program Interface
AQM	Active Queue Management
AR	Augmented Reality
ATN	Access Transport Network
BBR	Bottleneck Bandwidth and RTT
BGP	Board Gateway Protocol
BRAS	Broadband Remote Access Server
BRS	Burst Size
CDF	Cumulative Distribution Function
CIR	Committed Information Rate
CPU	Central Process Unit
CSFQ	Core-Stateless Fair Queuing
DCTCP	Data Center TCP
DHCP	Dynamic Host Configuration Protocol
DIP	Deterministic IP
DNS	Domain Name Service
DOS	Denial Of Service
DPI	Deep Packet Inspection
DSCP	Differentiated Services Code Point
Dst-EH	IPv6 Destination Extension Header
EH	IPv6 Extension Header or Extension Option
EPC	Evolved Packet Core
FI	Flow Identification
HbH-EH	IPv6 Hop-by-Hop Extension Header
HbH-EH-aware node	Network nodes that are configured to process the IPv6 Hop-by-Hop Extension Header
HOPOPT	Hop Option
HW	Hardware
IANA	Internet Assigned Numbers Authority
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
IP	Internet Protocol
IW	Initial Window
LDP	Label Distribution Protocol
MEC	Mobile Edge Computing
MPLS	Multi-Protocol Label Switching
MPTCP	Multi-Path TCP
MS	Multi-Segment
MSS	Multi-Segment Size
NPU	Network Process Unit
NSIS	Next Steps In Signaling
OAM	Operation And Management
OS	Operating System
PCC	Performance-oriented Congestion Control
PDN	Packet Data Network
PERC	Proactive Congestion Control Algorithm
PGW	PDN Gateway
PIE	Proportional Integral controller Enhanced
PIR	Peak Information Rate
PLR	Packet Loss Ratio
POC	Prove Of Concept
QoS	Quality of Service
RCP	Rate Control Protocol
RFC	Request for Comments
RMCAT	RTP Media Congestion Avoidance Techniques
RSVP	Resource Reservation Setup Protocol
RTCP	Real Time Control Protocol
RTP	Real-time Transport Protocol
RTT	Round Trip Time
SCTP	Stream Control Transmission Protocol
SIS	Service ID Size
SLA	Service Level Agreement
SP	Service Provider

SYN Synonym
TC-ACK TCP acknowledgement packet
TCP Transport Control Protocol
TM Traffic Management
TOR Top-Of-Rack
UDP User Datagram Protocol
VR Virtual Reality
WFQ Weighted Fair Queuing
WG Working Group
XCP eXplicit Control Protocol

4 Introduction

4.1 IP and Transport Technologies

This clause briefs the IP and transport protocol and technologies.

The traditional IP network can only provide the best-effort service. The transport layer (TCP/UDP) on top of IP is based on this fundamental character of IP network. The best-effort-only service has influenced the transport evolution for quite a long time, and results in some widely accepted concepts, assumptions and solutions, such as:

- The IP layer can ONLY provide the basic P2P (point to point) or P2MP (point to multi-point) end-to-end connectivity in Internet, but the connectivity is not reliable and does not guarantee any quality of service (QoS) to end-user or application, such as bandwidth, packet loss, latency, jitter, etc. Due to this fact, the transport layer or application will have its own control mechanism for congestion and flow to obtain the reliable and satisfactory service to cooperate with the under layer network quality.

- The transport layer assumes that the IP layer can only process all IP flows equally in the hardware since the best effort service is actually an un-differentiated service with maximized fairness [i.3]. The process includes scheduling, queuing and forwarding for all IP flows equally. Thus, the transport layer is supposed to behave nicely and friendly to make sure all flows will only obtain its own faired share of resource, and no one could consume more resource and no one could be starved.

Clause 4.2 briefs the analysis of current transport related technologies including TCP, UDP, DiffServ, IntServ, and MPLS. The major focus is TCP since it is the most widely used and the most complicated transport protocol.

4.2 TCP Solution Analysis

4.2.1 TCP Overview and Evolution

As a most popular and widely used transport technology, TCP is the most popular transport protocol in Internet. TCP traffic is actually dominating Internet from the birth of Internet. It is key to analyse TCP to get any conclusion for the current transport technology, and give any new proposal. This clause will brief the TCP, its variations and some key characteristics.

The major functionalities of TCP are flow control and congestion control.

The flow control is based on the sliding window algorithm. In each TCP segment, the receiver host specifies in the receive window field the amount of additionally received data (in bytes) that it is willing to buffer for the connection. The sending host can send only up to that amount of data before it will wait for an acknowledgment and window update from the receiving host.

The congestion control is the algorithm to prevent the hosts and network device fall into congestion state while trying to achieve the maximum throughput. There are many algorithm variations developed so far.

All congestion control will use some congestion detection scheme to detect the congestion state and adjust the rate of source to avoid the congestion.

No matter what congestion control algorithm is used, all classical TCP solutions are pursuing three targets [i.4]:

1) Higher efficiency in bandwidth utilization.
2) More fairness in bandwidth allocation.
3) Faster convergence to the equilibrium state.

Recently, with the growth of new TCP applications in data center, more and more solutions were proposed to solve buffer bloat, incast problems typically happened in data center. These solutions include DCTCP, PIE, CoDel, FQ-CoDel, etc. In addition to the three classical TCP targets mentioned above, these solutions have another target which is to **minimize the latency**.

4.2.2 TCP Solution Variants

There are many TCP variants and optimization solutions since TCP was introduced 40 years ago. Below lists the major TCP variants including typical classical solution and some contemporary solutions proposed recently:

- The classical solutions:
  - These solutions are implemented on host only. They use different congestion detection and inference mechanism, either based on packet loss, RTT or both, to dynamically adjust the TCP window to do the congestion control, such as: TCP-reno [i.5], TCP-vegas [i.6], TCP-cubic [i.7], TCP-compound [i.8], TIMELY [i.9], etc.
- The explicit rate solutions:
  - These solutions do not use the traditional black box mechanism executed at host to infer the TCP congestion status. Instead, they rely on the rate calculation on routers to notify host to adjust accordingly. Both network devices and hosts need to be changed in software and/or hardware. Typical solutions are: XCP [i.10], RCP [i.11].

NOTE: XCP and RCP are described for TCP here is referring to the scenario when XCP and RCP are used with TCP.

- The AQM solutions:
  - These solutions use AQM (Active Queue Management) techniques on routers to control the buffer size or queuing, thus control the congestion and minimize the latency indirectly. Both network devices and hosts may need to be changed in software and/or hardware. They include: DCTCP [i.12], PIE [i.13], CoDel [i.14], FQ-CoDel [i.15], etc.
- The new concept solutions:
  - Unlike above categories, the category of these solutions use completely new concepts and methods to either accurately calculate, or figure out the optimized rate and latency for TCP, such as: PERC [i.16], BBR [i.17], PCC [i.18], Fastpass [i.19], etc.

4.2.3 TCP Throughput Constraints

For the traditional TCP optimization solutions, the efficiency target is to obtain the high bandwidth utilization as much as possible to approach the link capacity. The link utilization is defined as the ratio of the total throughput of all TCP flows on a network device to the network bandwidth of all links.

For individual TCP flow, its actual throughput is not guaranteed at all. It depends on many factors, such as TCP algorithm used, the number of IP (including TCP, UDP and all other type of IP protocols) flows sharing the same link, host CPU power, network device congestion status, physical propagation delay in transmission, etc.