ETSI GR NIN 001 V1.1.1 (2021-03)

Non-IP Networking (NIN); Problem statement: networking with TCP/IP in the 2020s

Disclaimer

The present document has been produced and approved by the Non-IP Networking 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/NIN-001

Keywords
autonomic networking, core network, fixed networks, intelligence-defined network, internet, layer 3, network monitoring, network performance, network scenarios, next generation protocol, non 3GPP access

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 prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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

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) Non-IP Networking (NIN).

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.

Introduction

The TCP/IP suite of network protocols is now over 40 years old and was designed for different requirements than the networking of the 2020s. This raises addressing, mobility, performance, and security issues that have required significant effort, energy, and cost to mitigate, and have been well documented, for instance in ISO/IEC TR 29181-1 [i.1].

Any form of wireless comms, whether they be 2/3/4/5G, satellite or Wi-Fi, needs to go through a 'wired' back-haul at some point, if not multiple points. How do all these technologies converge and connect edges to the fixed 'backbone'?

Connecting across the world on an architecture where ageing network protocols (IP, MPLS and more) are not able to move beyond best effort networking makes it difficult to meet the increasingly challenging guaranteed end-to-end SLAs required for high value, business- and safety-critical applications. National and international core networks need to work in conjunction with future internet protocol paradigms required to satisfy future network demands and deliver the promised 5G benefits, while being secure, robust, trusted, and resilient by design.

With the increasing challenges placed on modern networks to support new use cases (some of which require ultra-low latency) and greater connectivity, Service Providers are looking for candidate technologies that may serve their needs better than the TCP/IP-based networking used in current systems.

1 Scope

The present document describes the challenges of IP-based networking for fixed and mobile networks and ways in which new network protocols can result in improved performance and more efficient operation. Topics covered include:

- efficient use of spectrum;
- efficient forwarding;
- naming and addressing (including addressing lifecycle);
- mobility and multihoming;
- Quality of Service (QoS);
- time-sensitive networking;
- performance;
- authenticity, integrity, confidentiality, access control, and identifiers;
- lawful interception;
- ease of management; and
- migration from current technology.

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] ISO/IEC TR 29181-1:2012: "Information technology -- Future Network -- Problem statement and requirements -- Part 1: Overall aspects".

[i.2] T-Mobile/IoT - Whitepaper, March 019: "The Game Changer for the internet of things".

NOTE: Available at https://www.t-mobile.com/content/dam/tfb/pdf/Whitepaper-Narrow-BandIo-T2019.pdf.

[i.3] IETF RFC 1144: "Compressing TCP/IP Headers for Low-Speed Serial Links".

[i.4] IETF RFC 2508: "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links".

[i.5] ETSI TR 103 369: "CYBER; Design requirements ecosystem".

[i.6] ETSI TS 101 158: "Telecommunications security; Lawful Interception (LI); Requirements for network functions".

[i.7] ETSI TS 101 331: "Lawful Interception (LI); Requirements of Law Enforcement Agencies".

[i.8] IETF RFC 2205: "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification".

[i.9] IETF RFC 8578: "Deterministic Networking Use Cases".

[i.10] IEEE 802.1AS™: "IEEE Standard for Local and Metropolitan Area Networks -- Timing and Synchronization for Time-Sensitive Applications".

[i.11] IEEE 802.1Q™: "IEEE Standard for Local and Metropolitan Area Networks -- Bridges and Bridged Networks".

[i.12] Dr. N. Davies: "The properties and mathematics of data transport quality", 2009.

NOTE: Available at https://www.slideshare.net/mgeddes/intro-dataqualityattenuation.

[i.13] I. Johansson: "Congestion control for 4G and 5G access", Internet Engineering Task Force Internet Draft, July 2016.

NOTE: Available at https://tools.ietf.org/html/draft-johansson-cc-for-4g-5g-02.

[i.14] ETSI TS 133 210: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Network Domain Security (NDS); IP network layer security (3GPP TS 33.210)".

NOTE: Available at https://www.etsi.org/deliver/etsi_ts/133200_133299/133210/.

[i.15] "Patterns in network architecture: a return to fundamentals", chapter 6 'Divining Layers', J. Day, Pearson, 2008, ISBN 0-13-225242-2.

[i.16] P. Teymoori, M. Welzly, S. Gjessingz, E. Grasa, R. Riggio, K. Rauschk, D Siracusa: "Congestion Control in the Recursive InterNetworking Architecture (RINA)", IEEE ICC 2016 - Next-Generation Networking and Internet Symposium, 2016.

[i.17] GSMA, June 2019: "NB-IoT Deployment Guide to Basic Feature set Requirements".

NOTE: Available at https://www.gsma.com/iot/wp-content/uploads/2019/07/201906-GSMA-NB-IoT-Deployment-Guide-v3.pdf.

[i.18] IETF RFC 7426: "Software-Defined Networking (SDN): Layers and Architecture Terminology".

[i.19] IEEE Std 802™: "IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture".

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

QUIC: UDP-based transport and session-control protocol with claimed performance improvements over TLS/TCP

3.2 Symbols

Void.

3.3 Abbreviations

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

API	Application Program Interface
ARP	Address Resolution Protocol
AS	Autonomous System
BGP	Border Gateway Protocol
BLER	Block Error Rate
CA	Certificate Authority
CBOR	Concise Binary Object Representation
DHCP	Dynamic Host Configuration Protocol
DNS	Directory Name Service
HTTP	HperText Transfer Protocol
HTTPS	HperText Transfer Protocol Secure
IANA	Internet Assigned Numbers Authority
ICMP	Internet Control Message Protocol
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
IPX	Internetwork Packet Exchange
JSON	JavaScript Object Notation
LAN	Local Area Network
LI	Lawful Interception
LPWAN	Low Power Wide Area Networks
MAC	Media Access Control
MIMO	Multiple Input and Multiple Output
MPLS	Multi-Protocol Label Switching
NAT	Network Address Translation
NR	New Radio
OSI	Open Systems Interconnection
OUI	Organizationally Unique Identifier
PST	Pacific Standard Time
QoS	Quality of Service
RFC	Request For Comment
RIR	Regional Internet Registry
ROA	Route Origin Authorization
ROHC	RObust Header Compression
RPKI	Resource Public Key Infrastructure
RSVP	ReSource Reservation Protocol
RTP	Real-time Transport Protocol
SDN	Software-Defined Networking
SINR	Signal to Interference and Noise Ratio
SIP	Session Initiation Protocol
SYN	SYNchronize (TCP control flag)
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TSN	Time-Sensitive Networking
TTI	Time Transmission Interval
UDP	User Datagram Protocol
USA	United States of America
VPN	Virtual Private Network

4 Efficient use of cellular radio spectrum

4.1 Introduction

4.1.0 Requirement for efficiency

Radio spectrum is regulated, finite, and expensive for a cellular radio network to acquire and use. Spectrum is shared among all devices attached to the radio network to enable download (from the network to the device, the 'downlink') and upload (from the device to the network, the 'uplink'). If the operator can share the spectrum efficiently, among users and their applications, then the operator may reduce their costs, transmit more data in shorter time intervals, and enable more devices to communicate simultaneously. This clause summarizes existing techniques to use spectrum efficiently at the radio and other layers, and how non-IP network protocols could further improve spectral efficiency.

4.1.1 Definition of efficiency

In the present document, networking efficiency is defined as the number of application bits per Hz per second. Application bits are the data communicated between the client and server applications once all network headers have been removed.

4.2 Radio techniques for spectral efficiency

4.2.1 Space division

Radio frequencies experience path loss as they travel through space, proportional to the distance they travel and affected by environmental conditions (absorption losses) along the path. Eventually the signal is attenuated to the point where the Signal to Interference and Noise Ratio (SINR) is too low to carry information reliably. Cellular networks calculate the path loss to determine the boundaries of each cell - after which the frequency ceases to be reliable, assuming a fixed power level at the transmitter. This allows networks to reuse that same frequency in other cells, although to avoid interference at the cell boundary, the same frequency is typically not used in adjacent cells that abut that boundary.

Whilst traditional cellular antennae transmit signals in an arc to fill a portion of the cell, Massive Multiple Input and Multiple Output (MIMO) antenna systems allow the narrow targeting of signals - 'beamforming' - which provides higher throughput and reduced latency between the MIMO antenna and receiving device.

Spatial multiplexing allows more than one data signal to be transmitted and received simultaneously on the same channel. MIMO achieves this by utilizing multipath-propagation to increase the number of paths a signal can take from transmitter to receiver.

4.2.2 Time division

The same frequency may be reused to communicate information if the frequency is divided into timeslots. Attached devices are allocated a timeslot by the radio network and transmit or receive within that slot. The longer the timeslot, the more information can be communicated within it - but at the cost of reducing the number of devices that can be served on that frequency in a given time.

4.2.3 Efficient modulation schemes

Cellular networks from 2G onwards transmit digital data. The transmitter takes digital data as an input, and transmits analogue radio in a way that allows the receiver to parse and reconstruct the digital data. This is achieved through modulation - wherein characteristics of the analogue radio wave are controlled and adapted to signal information to the receiver - with both sender and receiver utilizing modulators-demodulators (modems).

Given that the speed of a radio wave is fixed - defined by the speed of light through air - there are three characteristics of a radio wave that can be modulated: amplitude, frequency and phase.

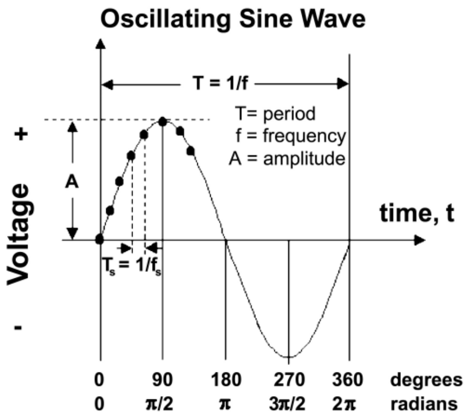

Figure 4.1: Characteristics of a sinusoidal wave (source: National Institute of Standards & Technology, 2010 Public Domain as per https://commons.wikimedia.org/wiki/File:Oscillating_sine_wave.gif)

When a device attaches to the cellular network, it is allocated a frequency from the operator's spectrum. Data to be communicated to/from the device is encoded into an analogue signal. One of the signal characteristics - amplitude, frequency, phase - is selected, and used to modulate a number of 'sub-carriers' (sub-divisions of the operator spectrum).

The resulting information is transmitted to the receiver, which demodulates the sub-carriers to retrieve the information.

The goal of a modulation scheme is to balance data throughput - the amount of data that can be modulated per second per Hz - with resilience from interference between the tightly-spaced frequency sub-carriers.

4.2.4 Radio resource block allocation

Devices continuously signal their received signal strength to the network. The network uses these values, as well as the size of packet queue for that user, to optimize the amount of radio resource to dedicate to that user, in the form of radio resource blocks. 5G New Radio (NR) introduces the concept of bandwidth parts, which allows the spectrum to be flexibly sliced up into groups of resource blocks for different users depending upon their needs: for example small groups for NB-IoT and large groups for enhanced mobile broadband.

4.2.5 Transmission interval

The Time Transmission Interval (TTI) is the duration of a transmission. A shorter TTI allows more transmissions per second, but with a reduced data payload (i.e. fewer radio resource blocks). The TTI can be adapted based on the application, for example a short TTI for Ultra-Reliable Low-Latency Communications with small. frequent payloads.

4.2.6 Narrowband IoT and Non-IP Data Delivery

Narrowband-IoT (NB-IoT) [i.17] is used for communication with low-power, low-throughput devices. The use of a narrow frequency band (200 kHz) at a lower carrier frequency reduces maximum throughput but also allows for signal penetration indoors.

NB-IoT may operate in a 'Non-IP Data Delivery' mode to remove the IP header from the transmitted payload, improving efficiency (see [i.2]). This is important as it minimizes radio transmissions, which apply a significant drain to the batteries of low-power devices.

4.3 Non-radio techniques

4.3.1 RObust Header Compression (ROHC)

IP encapsulation results in a per-packet header overhead: 20 bytes for IPv4, 40 bytes for IPv6 due to the increased address lengths. Transport protocols contribute an additional 20 bytes (TCP) or 8 bytes (UDP), and for "live" media such as audio there will also be at least 12 bytes of RTP header.

To tackle this problem, RObust Header Compression (ROHC, IETF RFC 1144 [i.3] and IETF RFC 2508 [i.4]) identifies redundant information (that which will be sent in every packet of a flow) and only transmits it in the first packet. Subsequent variable information (segment numbers, etc.) are transmitted in compressed form, and reassembled by the receiver.

Whilst ROHC may seem ideal to solve the problem of header overheads - reducing them to around 2 or 3 bytes - there are caveats:

- It incurs a cost to operators if they activate ROHC in a software licence.
- It incurs compute energy and latency.
- It requires a tuning of the Block Error Rate (BLER) used in radio transmission to ensure that the important information (the flow metadata) is not lost, since that would affect the following packets in the flow.

It may be for these reasons that operators typically only apply ROHC to VoLTE (Voice over LTE) flows. There are two reasons for this:

- the (IP, UDP, RTP) headers are almost the same size as the VoLTE payload (~60 bytes) , meaning header compression is more beneficial than e.g. video streams with per-packet payloads of 1 340 bytes;
- operators may be penalized by regulators if their voice call completion rate drops below a certain threshold, hence operators are more likely to invest in resources to ensure that does not happen.

Operators are less likely to apply ROHC to general Internet traffic due to the processing and licence costs. This is also the case with Low-Power IoT traffic, where the BLER issue is exacerbated. This has motivated new workarounds, such as Narrowband IoT (NB-IoT) omitting IP from the transmission entirely.

Rather than compressing and inflating IP and transport layer headers, ISG NIN proposes to tackle the problem at source through an efficient protocol design. This can reduce costs and latency, and avoid issues with Block Error Rates at transmission.

4.3.2 Payload compression

Payload - the application data carried in packets - is sized according to the amount of data and its serialization. Efficient serialization schemes reduce the overhead in representing data for interpretation by the receiving application: for example, JSON or CBOR (Concise Binary Object Representation). The entire data payload can be further compressed using gzip or similar for transmission with decompression at the client.

4.4 Areas that non-IP networking can help improve

4.4.1 Transmission overheads

Whilst ROHC (see clause 4.3.1) reduces IP header overheads significantly, it incurs financial cost, and requires compute and energy expenditure at both network and user equipment. Hence today it is typically used only for operator VoLTE (Voice over LTE) services, which have a significantly larger header-size to payload-size ratio than e.g. video streaming.

IoT services - especially Low Power Wide Area Networks (LPWAN) - are similar to VoLTE in that they have a large header-size to payload-size ratio (with a maximum throughput of 250 Kb/s on downlink). Removing redundant encapsulations and reducing header size will intrinsically improve networking efficiency for such services without a requirement for ROHC.

4.4.2 The propagation range of ultra-low latency services

The Transmission Time Interval (TTI) of the 5G radio air interface is ~140 μs. To realize this at cell edge in a network configured as described in clause 4.4.4, the radio payload per TTI can be no larger than ~18 bytes. ~100 byte payloads (TCP/IP + IPsec) will take up approximately 5 TTIs, which places a latency constraint on both the application and other applications competing for transmission blocks. Lower networking payloads means more efficient use of TTI.