International Telecommunication Union

ITU-T Technical Report

(2020)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

FG-NET2030 – Focus Group on Technologies for Network 2030

FG-NET2030-Gap

Network 2030 – Gap analysis of Network 2030 new services, capabilities and use cases

Summary

The Technical Report on gap analysis presents the current work on network and communication services that has been carried out by different standards development organizations (SDOs) with respect to the Network 2030 services, capabilities, and representative use cases. Based on these inputs, this report identifies gaps, namely issues and technologies that are not currently addressed, and will be required for the support of new use cases and the network infrastructure of 2030 and beyond. The report uses Network 2030 services and use case cluster reports as an input for the Network 2030 related requirements.

Keywords

Coordination guarantee, digital twin, haptic communication, holographic type communication, high precision communications, in-time guarantee, many-to-many, Network 2030, on-time guarantee, one-to-many, qualitative communication, qualitative communications.

Table of Contents

			Page
1	Scope		1
2	References		1
3	Definitions, abbreviations and acronyms		3
	3.1	Terms used in this technical report	3
	3.2	Abbreviations and acronyms	3
4	Network 2030 application delivery model		4
	4.1	Overview	4
	4.2	Methodology	5
	4.2.1	New services and capabilities Handles	6
	4.2.2	Network 2030 use case cluster handles	6
	4.2.3	Network 2030 Infrastructure and Operations Handles	7
5	Analysis of the network service model		7
	5.1	Description	7
	5.2	In-time and on-time service guarantees	7
	5.2.1	Baseline	7
	5.2.2	Gaps	8
	5.3	Coordinated services	9
	5.3.1	Baseline	9
	5.3.2	Gaps	9
	5.4	Qualitative communication services	9
	5.4.1	Baseline	10
	5.4.2	Gaps	10
	5.5	Haptic and tactile communications	10
	5.5.1	Baseline	10
	5.5.2	Gaps	10
	5.6	Very large volumetric-type communications (VLV) services	11
	5.6.1	Baseline	11
	5.6.2	Gaps	11
	5.7	Cut-through burst data forwarding	12
	5.7.1	Baseline	12
	5.7.2	Gaps	13
6	Analysis of network capabilities		13
	6.1	Description	13
	6.2	Network service interfaces	13
	6.2.1	Baseline	13
	6.2.2	Gaps	14
	6.3	High programmability and agile lifecycle	14
	6.3.1	Baseline	14
	6.3.2	Gaps	15
	6.4	Manageability	15

6.4.1 Assurance of high-precision services 15
6.4.2 Fulfilment and "operation-at-scale" 17
6.5 Security 18
6.5.1 Baseline 18
6.5.2 Gaps 18
6.6 Resilience 19
6.6.1 Baseline 19
6.6.2 Gaps 19
6.7 Loss-lessness 20
6.7.1 Baseline 20
6.7.1 Gaps 20
6.8 Privacy 21
6.8.1 Baseline 21
6.8.2 Gaps 21
6.9 Validation of delivered services 21
6.9.1 Baseline 22
6.9.2 Gaps 22
7 Analysis of network infrastructure and operations 22
7.1 Compute in networks 23
7.1.1 Baseline 23
7.1.2 Gaps 23
7.2 Intelligent operation networks 24
7.2.1 Baseline 24
7.2.2 Gaps 24
7.3 Support for ManyNets 25
7.3.1 Baseline 25
7.3.2 Gaps 26
7.4 Artificial intelligence aware networking 26
7.4.1 Baseline 26
7.4.2 Gaps 27
7.5 Support of SIoT-enabled logistics 28
7.5.1 Baseline 28
7.5.2 Gaps 28
8 Evaluation of key findings 29
8.1 Use cases to services and capabilities relevance 29
8.2 Factors considered for overall analysis 29
9 Summary and conclusion 32
Appendix I 33
Appendix II 36

ITU-T FG NET2030 Technical Report

Network 2030 – Gap analysis of Network 2030 new services, capabilities and use cases

1 Scope

This technical report presents the gaps in current network and communication technologies, which need to be addressed in order to meet the challenges introduced by future network applications in Network 2030. These gaps are intended to accommodate the use cases and meet the requirements that have been analysed by the ITU-T Focus Group on Network 2030 (FG-NET 2030).

The gaps presented are a summary of the detailed analysis already covered in Network 2030 services [4]. At a high-level, the goals are to identify the gaps between present 'Best-effort' to new 'Beyond best-effort' services. The end-to-end realization of new use cases requires not only new network services but also security, high-capacity, end-to-end service assurance, amongst others identified as essential capabilities.

The report also focuses on the major technical gaps of "Representative use cases and key network requirements for Network 2030" [2][3] and the analysis of how to satisfy them. The major challenges identified relate particularly to machine to machine communications, autonomous operations, specific bandwidth requirements, and the finest possible granularity of time-engineered services.

2 References

The following references contain provisions which, through reference in this text, constitute provisions of this report. At the time of publication, the editions indicated were valid.

[1] Network 2030 – A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond (May 2019).

[2] Technical Report: Representative use cases and key network requirements for Network 2030 (January 2020).

[3] FG NET-2030 Sub-G1 an Update of 2nd Report of Use cases and network requirements for Network 2030.

[4] New Services and Capabilities for Network 2030: Description, Technical Gap and Performance Target Analysis (October 2019).

[5] Time-Sensitive Networking (TSN) Task Group https://1.ieee802.org/tsn/ Referenced on 9 June 2020.

[6] Finn, N., Thubert, P., Varga, B., and J. Farkas, "Deterministic Networking Architecture", RFC 8655, October 2019, https://www.rfc-editor.org/info/rfc8655.

[7] A. Clemm, M. F. Zhani, R Boutaba: "Network Management 2030: Operations and Control of Network 2030 Services. https://doi.org/10.1007/s10922-020-09517-0 Journal of Network and System Management, Springer, March 2020.

[8] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," IEEE Pervasive Comput.,vol. 8, no. 4, pp. 14–23, Oct./Dec. 2009.

[9] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in Proc. ACM 1st Edition MCC Workshop Mobile Cloud Comput., Helsinki, Finland, 2012, pp. 13–16.

[10] European Telecommunications Standards Institute. MobileEdge Computing (MEC) Terminology. Accessed on May 2017. Available: http://www.etsi.org/deliver/etsi_gs/MEC/001_099/001/01.01.01_60/gs_MEC001v010101p.pdf

[11] European Telecommunications Standards Institute. Multi-Access Edge Computing. Accessed on May 2017. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/multi-accessedge-computing

[12] J. S. Preden et al., "The benefits of self-awareness and attention in fog and mist computing," Computer, vol. 48, no. 7, pp. 37–45, Jul. 2015.

[13] C. Baktir, A. Ozgovde and C. Ersoy, "How Can Edge Computing Benefit From Software-Defined Networking: A Survey, Use Cases, and Future Directions," in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2359-2391, Fourthquarter 2017.

[14] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. D. Foy, Y.Zhang, "Mobile Edge Cloud System: Architectures, Challenges, and Approaches," IEEE Systems Journal, pp. 1-14, Feb, 2017.

[15] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2322-2358, 2017.

[16] P. Mach, Z. Becvar, Mobile Edge Computing: A Survey on Architecture and Computation Offloading, IEEE Communications Surveys and Tutorials, vol. 19, no. 3, pp. 1628-1656, third quarter 2017.

[17] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," Wireless Commun. Mobile Comput., vol. 13, no. 18, pp. 1587–1611, 2013.

[18] Sterbenz, J. P. G., Hutchison, D., Çetinkaya, E. K., Jabbar, A., Rohrer, J. P., Schöller, M., & Smith, P. (2010). Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. Computer Networks, 54(8), 1245–1265.

[19] C. Filsfils, S. Previdi, L. Ginsberg, B. Decraene, S. Litkowski, R. Shakir: Segment Routing Architecture. IETF RFC 8402, July 2018.

[20] Intent-Based Networking – Concepts and Definition: https://www.ietf.org/id/draft-irtf-nmrg-ibn-concepts-definitions-01.txt

[21] Zero-touch network and Service Management (ZSM); Reference Architecture https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf

[22] P. Mach, Z. Becvar, (2017). Mobile edge computing: A survey on architecture and computation offloading. IEEE Communications Surveys & Tutorials, 19(3), 1628-1656.

[23] Chen, M., Hao, Y., Hu, L., Hossain, M. S., & Ghoneim, A. (2018). Edge-CoCaCo: Toward joint optimization of computation, caching, and communication on edge cloud. IEEE Wireless Communications, 25(3), 21-27.

[24] W. Sun, J. Liu, Y. Yue (2019). AI-enhanced offloading in edge computing: when machine learning meets industrial IoT. IEEE Network, 33(5), 68-74.

[25] L. Atzori, A. Iera, and G. Morabito. The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization. Computer Networks. 2012.

[26] L. Atzori, A. Iera, and G. Morabito. Sociocast: A new network primitive for the IoT. IEEE Communications Magazine. 2019.

3 Definitions, abbreviations and acronyms

3.1 Terms used in this technical report

This Technical Report uses terms and definitions found in [2], [3] and [4].

3.2 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

ABF	Application-aware Burst Forwarding
AIN	Artificial Intelligence aware Networking
AQM	Active Queue Management
BDF	Burst Data Forwarding
BIER	Bit-Indexed Explicit Replication
Caps	(Network 2030) Capabilities
CBR	Constant Bit Rate
CO	Compound Services
DDoS	Distributed Denial-of-Service
DETNET	Deterministic Networking
DL	Deep Learning
DT	Digital Twin
ECN	Early Congestion Notification
FO	Foundational Services
GTP	GPRS Tunnelling Protocol
HSD	Huge Scientific Data
HTC	Holographic Type Communications
HTCS	Holographic Type Communication Services
IBN	Intent Based Networking
IIoT	Industrial IoT cloudified
ION	Intelligent Operation Networking
ISL	Inter-Satellite Link
L4S	Low Latency, Low Loss, Scalable
LEO	Low-Earth Orbit
MTU	Maximum Transmission Unit
NCC	Network Compute Convergence
NFV	Network Function Virtualization
PIM	Protocol Independent Multicast
QoS	Quality of Service

QUIC	Quick UDP Internet Connections
RTT	Round-Trip Time
SC	Service Capabilities
SIoT	Social Internet of Things
SLO	Service Level Objective
STIN	Space-Terrestrial Integrated Network
TCP	Transmission Control Protocol
TIRO	Tactile Internet and Remote Operations
TSN	Time Sensitive Networking
UCC	Use Case Clustering
UDP	User Datagram Protocol
VBR	Variable Bit Rate
VLV	Very Large Volumetric-Type Communications

4 Network 2030 application delivery model

Network 2030 applications can be described in terms of the new use case clusters and emerging infrastructures as outlined in [2] and [3]. In order to understand a comprehensive solution towards Network 2030, we first analyse the gaps via a reference framework referred to as "the application delivery model for Network 2030 applications". It follows the current end-to-end communication model with the Network 2030 technology as shown in Figure 1. The different interconnecting endpoints are aware of, and use, the new services offered by the networks.

NOTE – This model is derived from the one described in [4].

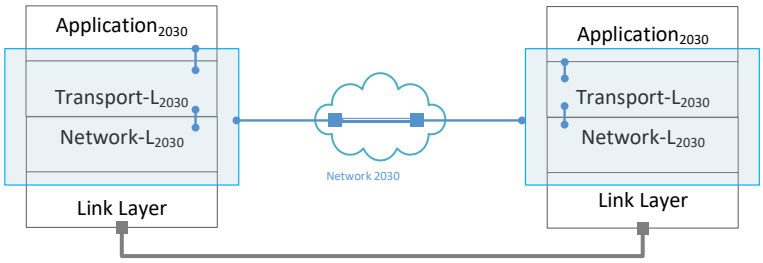

Figure 1 – Network 2030 enabled end to end communication model

4.1 Overview

A collective summary of the two studies can be described through a set of expectations which we envisage future applications and industry verticals will require from the networks in 2030, see Figure 2. Many applications of Network 2030 will utilize characteristics of one of the uc-clusters (use case clusters), viz, digital twin (DT), tactile Internet and remote operations (TIRO), industrial IoT cloudified (IIoT), application burst forwarding (ABF), huge scientific data (HSD) and holographic type communications (HTC). The group of forward looking use cases have specific requirements, different to those supported in today's networks.

The support for these uc-clusters is enabled through foundational (FO) services, compound (CO) services, and capabilities (Caps). These services do not exist in the networks today.

It is expected that networks in 2030 will be an integration of existing and emerging infrastructure such as space-terrestrial integrated network (STIN) and ManyNets. The mechanisms to describe technological, business and economic models for this new infrastructure will be different from those of today.

The uc-clustering specifically identified artificial intelligence aware networking (AIN) and network compute convergence (NCC) functional requirements in the networks. These functions apply to both the applications and operations of Network 2030. However, pervasive automation will also drive the overall operations and management of the networks themselves to be far more autonomous through intelligent operation networking (ION).

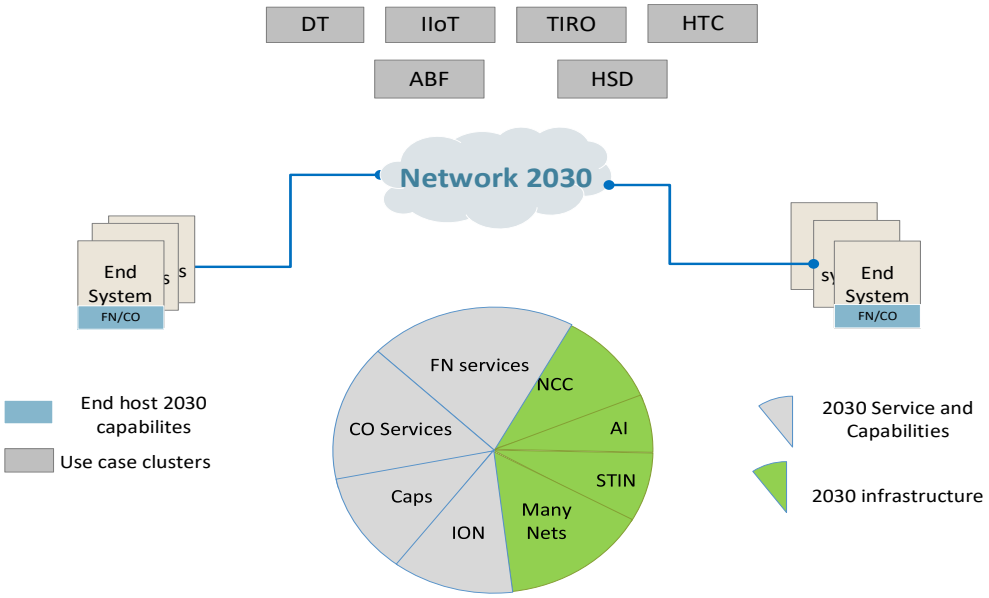

Figure 2 – Network 2030 use cases and services

The above discussion is illustrated in the Network 2030 reference model shown in Figure 1.

4.2 Methodology

This report is derived from a detailed stepwise gap analysis performed as part of the identification and description of new services for Network 2030 in [4]. Additional gaps emerged during the study of uc-clusters [2][3]. Hence, the overall gaps are categorised as follows:

1) **Gaps in network services**: services are functional entities that assist (or are used by) the end user applications to meet their requirements related to communicating with other end users or applications.

2) **Gaps in network capabilities**: are additional functions that facilitate operations of those services.

3) **Gaps in network compute**: converged ubiquitous compute that supports across all the networks.

4) **Gaps in intelligent operation network**: to address the new network requirements.

5) **Gaps in support of ManyNets**: that are new access and infrastructure changes influencing the networks of 2030.

6) **Gaps in artificial intelligence aware in networking**: identifies the use of AI in the operations of the networks.

7) **Gaps in social IoT**: new communication schemes for IoT applications.

NOTE – The gap analysis covers uc-clusters and services and capabilities (SC-group). Network architecture aspects are not covered in the scope of this Technical Report.

While preparing use cases and services, the requirements were highlighted in the context of the detailed descriptions. The scope of this report is to identify the gaps. However, FG-NET-2030 SubG-1 and SubG-2 have extensively researched new scenarios and a few novel concepts are introduced where baseline comparisons are not possible. In these cases the requirements are stated instead.

4.2.1 New services and capabilities Handles

Network 2030 services [4] are defined as new network-layer services in the data plane. These services are complemented by the network capabilities necessary for both the execution of services and the corresponding operations in the networks. The services and capabilities are labelled as listed in Table 1.

Table 1 – Network 2030 services and capabilities labels

	Relevant network requirements
HPC	In-time, On-time services
COOR	Coordinated services
QUAL	Qualitative services
TAC	Tactile network service
VLV	Very large volumetric type communications service
NSI	Network service interfaces
PRG	High programmability
MGM	Manageability
SEC	Security
RES	Resilience
LLS	Loss-lessness
PRIV	Privacy
VDS	Validation of delivered services
BDF	Cut-through burst data forwarding

4.2.2 Network 2030 use case cluster handles

The study of the uc-clusters in [2] and [3] identified emerging network characteristics and the corresponding gaps in need of analysis. Use case clusters themselves are described with respect to basic network resources, which in turn may be categorized through services. Table 2 lists use case cluster labels.

Table 2 – Use case cluster labels

Cluster Id	**Relevant network requirements**	**Service coverage**
HTC	Holographic type communications (HTC)	VLV, QUAL, COOR
TIRO	Tactile Internet for remote operations (TIRO)	TAC, IOTS, LLS
ION	Intelligent operation networking (ION)	PRG, MGM, SEC, RES, PRIV, VDS
NCC	Network and computing convergence (NCC)	PRG, MGM, SEC, RES, PRIV, AI, NSI

Table 2 – Use case cluster labels

Cluster Id	Relevant network requirements	Service coverage
DT	Digital Twin (DT)	TAC, VLV
STIN	Space-terrestrial integrated network (STIN)	MNET
IIoT	Industrial IoT (IIoT) with cloudification	IOTS, TAC, LLS
ABF	Application burst forwarding	BDF
HSD	Huge scientific data use case cluster	VLV, NCC
SIoT	Social IoT	PRG, MGM, SEC, RES, PRIV, NSI

4.2.3 Network 2030 Infrastructure and Operations Handles

The uc-clusters in [2] and [3] also identified emerging network characteristics and corresponding gaps requiring analysis. Table 3 lists network operations characteristics.

Table 3 – Network operations characteristics

Infrastructure ID	Use case cluster name
MNETS	ManyNets
AI	Artificial intelligence
NCC	Network and compute convergence

As per clause 4.2.2, these use case clusters are also described with respect to basic network resources, which in turn may be categorized through services.

5 Analysis of the network service model

5.1 Description

The Network 2030 foundational services provide guarantees of specified time characteristics for in-time services, on-time services (HPC), and coordinated services (COORD). The goal of qualitative services (QUAL) is to associate a certain quality or value within the payload and accordingly treat parts of the payload based on those attributes.

This clause relates to clauses 6.1.3, 6.2.3 and 6.3.3 in [4].

The in-time and on-time services are characterized by specific guarantees of time; however, a certain expectation of data rate guarantee may also be associated with these services. Current means of providing low latency or bandwidth are covered through Internet QoS, TSN and DetNet.

5.2 In-time and on-time service guarantees

In-time services require that packet delivery latency across a network must not be exceeded. On-time services operate at a fine-grained timescale granularity (e.g., microseconds) and should be able to offer deterministic latencies with an extremely small window of bounded arrival times.

5.2.1 Baseline

Delivery of services with different objectives and requirements are achieved through the IP based quality of service (QoS) mechanisms such as integrated services (IntServ) and sifferentiated services (Diffserv).

DiffServ is a multiplexing technique used to manage bandwidth for different classes of traffic. The Diffserv markings are treated uniformly with in a network domain but the implementations across the network nodes can be different. In machine-type communications, end-to-end delivery times for the packets within a flow or group of flows could also vary and the QoS classes do not serve as an accurate indicator of delivery times.

The IntServ is a guaranteed service that provides per-flow fixed bandwidth guarantees and is based on the concept of reserving resources in advance for a given flow and are not available for other flows. IntServ traffic is shaped at the ingress network to ensure it does not exceed its consumption beyond what has been reserved. This must be done for each hop to support latency guarantees.

Time-sensitive networking (TSN) [5] is a suite of protocols that specifies scheduling mechanisms and time synchronization standards over a layer 2 network. Thus, TSN is not routing-capable to support a high density of endpoints. TSN does not aim to provide an on-time guaranteed service over large-scale networks nor over longer distances. Like IntServ, TSN is geared towards constant bit rate (CBR) traffic, not variable bit rate (VBR) traffic, and does not support the slowing down of packets based upon the earliest required delivery time.

The deterministic networking architecture (DetNet) [6] provides per-flow service guarantees in terms of (1) the maximum end-to-end latency (called bounded delay in DetNet) and bounded jitter, (2) packet loss ratio, and (3) an upper bound on out-of-order packet delivery. The DetNet keeps track of the per-flow state to implement advanced traffic shaping and packet scheduling schemes at every hop. This is not scalable because core routers can receive millions of flows simultaneously.

5.2.2 Gaps

G.NS.1. **Quantifiable end-to-end latency**

These services require that packet delivery latency across a network must not be exceeded (in-time), or operate at fine-granularity timescales (on-time). None of the current low-latency technologies support discrete upper and lower bounds for desired latency. The knowledge of exact latency values will allow schedulers to satisfy low latency requirements of multiple applications at the same time. Development of new algorithms or enhancements of several existing schedulers (as in TSN) are topics for further study.

G.NS.2. **Service constraints with varying bit rates**

In packet switched networks there are very few CBR services. Most data communication is interactive and data rates vary, therefore, bandwidth reservation techniques such as IntServ do not scale. Several examples in the use case clusters in clause 4.2.2 (HTC and IIOT) do not assume CBR traffic, in fact real time immersive media applications will require transmission at not-constant bit rates. Support for VBR traffic with on-time guarantees is necessary, with the constraint that it will not exceed the specified packet rate.

G.NS.3. **Application defined service customizations**

No two machine-type applications have the same parameters for in-time guarantees. One of the reasons to offer high-precision functionality as a service is for the applications to customize them according to specific needs. Therefore, the constraints mentioned above should be on a per packet basis since for in-time services, statistical flow shaping, and scheduling will lead to violation of the service for several packets in a flow. (For example, if the packet transmit timing was controlled from the sender, and an intermediate network node experienced congestion, it itself may end up sending packets). There is also a preference to trigger these customization requests from source endpoints for accurate end-to-end guarantees. Such an interface is missing and is further discussed in clause 6.2.

G.NS.4. **Pacing and queuing based on desired latency**

To utilize networks better with in-time services, scheduling of packets based on the desired earliest delivery time should be performed. The best-effort queuing disciplines cannot prioritize packets based on their desired end-to-end latency; at times it may need to slow down a packet in transit. While TSN performs in-time packet deliveries with strict time synchronization, as the network size grows its throughput deteriorates.

5.3 Coordinated services

A coordinated service provides a guarantee of delivery of multiple flows based on inter-dependencies between flows (known as co-flows), with respect to time-based, ordering, sequencing, and QoS rate being shared.

5.3.1 Baseline

Coordinated services relate to group communications such as multicast communications but are more than that. They involve coordination of more than one flow between the same source and destination, or flows between multiple endpoints. The multicast protocols such as protocol independent multicast (PIM) must find all the members that are part of the group G and forward data along a multicast tree from the multicast source S. The forwarding multicast routers need to maintain trees for each of the paths (S and G). Bit-indexed explicit replication (BIER) is proposed to eliminate group-specific signalling and complexity introduced from building multicast trees in the network is moved to the source. Bit to IP address resolution is required in BIER. Alternately, over-the-top (OTT) server-based group communications is another option where multiple users are connected to a server. The server sends a unicast message to each group member. In both cases, membership of the group must be maintained. The multicast routing protocols provide in-network group communications using source- and receiver-based trees while those that are server-based are n^2 peer to peer connections.

5.3.2 Gaps

G.NS.5. **In-network support for coordination**

The IP multicast set of protocols neither carry any dependency information nor actively performs coordination. This prevents exchange of information about inter-dependency among the co-flows, thus the data arrives at different receivers without synchronization. This type of dependency is required for DT and IIoT type use cases.

G.NS.6. **Application defined dependency**

The type of dependency required between different members is known to the application and the means to translate those into network functionality do not exist. Same as G.NS.3.

G.NS.7. **Support for dynamic pacing and feedback**

Coordinated services are bi-directional, interactive and often real-time communications. Therefore, adapting to the dynamic changes inside the network is critical and may be served better by having both state and co-dependency carried with the packet instead of maintaining state in the routers. Current protocols in particular favour shortest paths, unaware of coordination pace of packet or group of packet deliveries. This may require further buffering capabilities in the routers.

5.4 Qualitative communication services

The qualitative communication services associate semantics or a certain value to the user payload. This allows the payload to be serviced not as a single raw stream of bits, but by differentiating relevance or semantics into different chunks within the payload.

5.4.1 Baseline

In packet-switched networks, packets are the basic unit of transmission and are subjected to in-network treatment for forwarding, classification, discard, QoS, etc., as entirely one unit. As a result, network protocols have evolved to ensure that user data coded in a series of bits (0s and 1s) by the sending entity is exactly equal, bit-by-bit, at the receiving end. Under normal network conditions this is ideal, but can worsen network utilization in congested scenarios. Adaptive measures to network conditions occur only after the sender has learnt about end-to-end round-trip times (RTTs). For example, media streaming applications such as MPEG-DASH take large media streams and break them down into chunks with different attributes of time, resolution, etc. before transmitting.

Qualitative communication is a new service and works at sub-payload level. Intuitively, qualitative services attach different attributes within the payload and provide a means to drop, repair and recover user data in the network without compromising data integrity. Network coding technology can also recover and repair, as it allows intermediate nodes to generate new packets from incoming packets which can be decoded at the receiving nodes. Network coding works at the packet level, i.e., operates on payload and header collectively.

5.4.2 Gaps

G.NS.8. **Semantic- and context-based payload realization**

These two mechanisms describe functions of reliability, robustness and packet loss that qualitative services aim to improve upon by applying them at smaller granularity upon qualitative payloads. In order to differentiate the byte-stream in payloads from the meaningful or useful information, research is required to generate semantics and context awareness in the payload.

G.NS.9. **Application level packetization and encoding for qualitative payloads**

The characterization and the degree of significance of qualitative information is decided and assigned by the source application. The context should then be attached to chunks in the packet and shared with the network to operate on a qualitative packet without needing to look inside the payload. This is a new way to packetize data from applications but the ability to associate significance or context is not possible.

G.NS.10. **Forwarding-node qualitative function**

This is a new service requirement and network forwarding nodes need to perform new types of packet editing operations where chunks with lower significance are dropped from a packet upon congestion while retaining as much information as possible. Several new qualitative context aware functions will be devised to support such services.

5.5 Haptic and tactile communications

The haptic or tactile networking services are defined as composite services that require round-trip haptic feedback with known latency. Critical applications within these communications require high reliability.

5.5.1 Baseline

Haptic and tactile communications are an emerging field and do not have a packet-based transmission standard or well-established means of communication. No network level work has been carried out, except the establishment of dedicated physical links between remote sites.

5.5.2 Gaps

Since this is a composite service, its performance targets are translated from specific gaps in other services and capabilities. For example, ultra-low latency gaps are identified under clause 5.2.2; ultra-low packet loss, in clause 6.8.2; and, ultra-high bandwidth in clause 5.6.2. Additionally, haptic

use cases often involve multiple streams, (visual, touch and audio), each with varying bit-rates and levels of reliability.

G.NS.11. **Paced synchronization and prioritization**

By identifying mechanisms for the coordination of multiple flows for the same destination, synchronization as described in clause 5.3.2 is achieved. Coordinated communications with priority as a characteristic of co-flows.

5.6 Very large volumetric-type communications (VLV) services

The large volumetric type of communication services related gaps are generalized from the study of:

a) Holographic-type communication services [4] required by use case cluster (HTC)

b) Huge scientific data use case cluster (HSD).

5.6.1 Baseline

Current multimedia streaming services can support scales of 4K media over the Internet and a degradation in streaming quality over the network is generally endured because of the non-engaging nature of current media streams. On the other hand, networked AR/VR media formats are still under study. For such applications, providing efficiency in transport protocols will be more important than bandwidth with high capacity. High resolution or multi-view media applications use very high compression encoding schemes to reduce the transmitting bandwidth. The trade-off between delays due to compute time of compression and volume of transmission are being studied for real time network AR/VR. There are no definitive solutions yet because packet loss (and hence throughput) in the network becomes unpredictable. The details of HTC are covered in the HTC-cluster and demonstrate that resource requirements are even higher than AR/VR.

In terms of end-to-end data transmission, the bulk of packet switched applications use transmission control protocol (TCP) as their transport. It is anticipated that quick UDP Internet connections (QUIC) will become the new transport protocol. It is designed to bring a modular approach to transport related features, with a focus on security and fast session setup. Additionally, several new features are in the works (multi path, low latency, etc.) but fundamentally QUIC remains like TCP as an end to end transport. In-network support, transport protocols rely on early congestion notification (ECN). The L4S framework utilizes multiple features viz. ECT classification and low-latency queueing with active queue management (AQM) to generate in-network low queuing delays.

5.6.2 Gaps

The main gaps are related to the ability of networks to support foundational services with sufficiently low latency (in-time services with quantifiable latency) and sufficiently high bandwidth. Combining these two characteristics are extremely challenging with respect to the current structure of transport protocols.

G.NS.12. **Application defined consistent throughput**

In end-to-end transport control, throughput is determined by a packet-loss ratio equation. In HTC, packet loss can degrade user experience significantly. This is due to the fact that if a packet drops in a high data rate media transmission occurs, retransmission is required, and streaming has to wait for the retransmitted packets to arrive. Today's networks do not support predictable throughput (i.e., keeping the probability of packet loss constant), however the L4S PI^2 algorithm aims to minimize packet losses tremendously for low latency traffic. It focuses on fairness of throughput for the queue but still does not discriminate between different flows. i.e., from a network node's queue perspective, it is not obvious which packet from which flow was dropped. If the network and end

users were to support qualitative communications, then another dimension alongside qualitative payloads could be used to reduce packet drops.

G.NS.13. **Metadata for in-network transport feedback**

Delay and loss-based protocols that run end-to-end transport for better management of congestion are usually unaware of the detailed network conditions. Throughput and latency demands need to be communicated, at a higher granularity than DiffServ, etc. to the network, to ensure differentiated 'in-network' treatment where guarantees are necessary. For example, consider that a flow requires a certain amount of bandwidth guarantee that is not possible on the shortest path. Then it may be sufficient to provide the network with such metadata that flags this requirement; but the path the network uses to achieve the guarantee should be transparent to the applications. This can be contrasted against current developments that provide path segments from the source.

G.NS.14. **On demand, high-volume resource customization**

In a distributed workflow system, the loss of one node affects the whole system, so each node needs an end-to-end guarantee. At the same time, different research applications require different scales of bandwidth, which can be minutes, days, or long-term. The network should have the ability to dynamically provide bandwidth and share resources for effective utilization. End-to-end dedicated bandwidth guarantees need to preempt background traffic as well as the transfer of scientific data.

G.NS.15. **Synchronization from multiple sources**

Many scientific devices usually collect and transmit continuous data streams collected by multiple sources (e.g., radio telescopes distributed at different locations) to a remote processing centre for real-time analysis during observation (see HSD use case). The delay of one node's flow will result in the delay of the analysis and results.

This gap for HSD can utilize coordinated communication services where multiple sources form co-flows to exchange data and results with the processing centre. This differs from G.NS.14 because in this scenario raw scientific data is streamed from the sources. Whereas in the previous case, the data has already been collected and stored from previous experiments.

G.NS.16. **Reliable high-speed data streaming**

The local storage sizes of scientific applications are usually small or even non-existent, making it difficult to re-transmit lossy data and therefore traditional congestion and flow control algorithms do not apply. Reliability in this context is therefore concerned: (a) with no-congestion throughput maintenance at the transport level, and (b) the utilization of lower level link reliability parameters which requires coordination between upper and lower layers.

5.7 Cut-through burst data forwarding

Applications requiring short-duration, high volume data transmissions will utilize this service. Such services are required to handle data-bursts when the network node does not have enough bandwidth to handle burst forwarding. This network service aims to provide the same or identical burst rates at both the receiver and sender side.

5.7.1 Baseline

This service is derived from the application aware data burst forwarding (ABF) use case described in the use case report [3].

In general, a network forwards big data chunks as a byte stream or multiple correlated byte streams. The data generated by an application, which is usually larger than the maximum transmission unit (MTU), typically 1.5KB, is segmented and encapsulated into many MTU-sized packets. The host side distributed congestion control algorithms (with TCP or QUIC) are designed to equally share the congestion link bandwidth between different concurrent flows. The TCP provides self-adaptive throughput control (QUIC provides equivalent functionality on top of UDP in the user space).

Other alternatives, e.g., MPTCP and SCTP provide multiple streams to increase the throughput and resilience between two endpoints. None of these options are suitable for the period of cut-through bursts due to fairness among the concurrent flows.

5.7.2 Gaps

When forwarded over a bandwidth converged network, the completion of application data-burst usually takes longer than the data processing time at the receiver side because the entire application data needs to be received to start processing. The resulting lengthy data transmission time is due to low compute resource utilization. This is not desirable for cut-through bursts (no in-network queuing delays) in applications described in the ABF cluster.

G.NS.17. **Cut-through application data forwarding with dynamic link resource**

To minimize the end to end data transmission delay, the network should keep scheduling the data packets that belong to the same application data unit. This requires the data to be tagged so that the network can identify which application data should be forwarded using cut through, i.e., the network should transmit the packets from an application-data burst at the same speed as it is injected into the network.

However, such cut through technology requires uniform physical end to end bandwidth. An application expects the burst forwarding network to support dynamic (short duration) end to end virtual channel creation. A virtual channel must meet the bandwidth requirement of the current transmitting burst. After the burst is forwarded, the created virtual channels should be pulled down immediately. Alternately, these applications may utilize coordinated services with all the packets in the application-data burst tagged as co-dependent with the specified data-rate. The network will still be responsible for making the cut-through bandwidth available.

G.NS.18. **Congestion free cut-through burst transmission scheduling**

Uncoordinated burst forwarding in a bandwidth converged network can cause in-cast problems. To avoid network congestion, each burst transmission needs to be carefully arranged. Different from the current *host based distributed congestion control* algorithms, the burst transmission management algorithm should work as an on/off switch for burst transmission scheduling. Once the burst transmission is granted, the burst is sent using the line rate from the data source. If multiple data sources want to send data concurrently, the algorithm needs to guarantee that the data injected into the network does not produce congestion that overflows the available router buffer.

6 Analysis of network capabilities

6.1 Description

The available set of network capabilities are well-known and are being extensively used. As we transition to Network 2030 applications, these capabilities will need to evolve accordingly. The considerations provided here are specific in relation to Network 2030 and do not concern any generic evolution occurring regardless of Network 2030 capabilities.

6.2 Network service interfaces

6.2.1 Baseline

As stated in clause 8.1 of [4], network service interfaces (NSIs) will take an evolutionary approach and support well-known interface patterns. The most common pattern involves the use of sockets which allow a sender to write and a receiver to read.

At the same time, network service interfaces will need to allow applications to access and take advantage of Network 2030 services, requiring an evolution of interfaces to provide support for e.g., high-precision services.

While sockets are end-to-end, the gaps highlighted below extend to user network interfaces (UNI). The following clauses cover both application-to-network and end-to-end gaps.

6.2.2 Gaps

The following gaps need to be addressed for in-time and on-time services:

G.CAP.1. **SLO-aware network service interfaces**

Network service interfaces need to allow senders to specify service level objectives (SLOs) that are required to meet end-to-end latency targets, for in-time and on-time services. For example, this could involve the ability to indicate an SLO as a parameter when a socket is opened or created, or the ability to select between different socket types that are associated with different SLOs.

G.CAP.2. **SLO negotiation via network service interfaces**

Facilities to negotiate SLO targets need to be provided between application and network. For example, senders need to be provided with an indication whether a requested target will indeed be supported and with ways that permit "negotiation downwards" to identify a target that can be supported, say by allowing the NSI to provide an alternative "counter offer".

G.CAP.3. **Support for coordinated services**

Network interfaces are an interface between applications and the network. Therefore, applications that depend on coordinated services will need an interface to specify groupings of interdependent flows and the nature of the interdependency. Interfaces also need to manage the coflows dynamically. For example, change in either coflows membership or interdependent parameters.

G.CAP.4. **Support for qualitative services**

Support for qualitative communications is entirely application dependent. An application specifies and builds the context of a payload to differentiate between portions of it. Currently, sockets do not offer this type of differentiated payload structure. Likewise, NSIs should provide receivers with an indication if there were payload portions that could not be delivered.

G.CAP.5. **Inherent accounting support for SLOs**

SLO-aware accounting refers to the ability to validate whether service level objectives have been met. A separate infrastructure for accounting and for validation of adherence to SLOs is required. High-precision services are premium services and not best-effort but "guaranteed", likely leading to the requirement to allow senders and receivers to get indications that service level objectives are being adhered to as part of the network service interface. Such facilities could include warnings (or an indication) that a specified in-time or on-time target has been compromised, or a "receipt" at the end of a session providing a summary of session level SLO accounting.

G.CAP.6. **Accommodation of network-level trust mechanisms**

NSIs need to accommodate trust mechanisms, for example to authenticate a sender where required by a service. Currently, authentication is handled at the application level and not supported by network service interfaces.

6.3 High programmability and agile lifecycle

6.3.1 Baseline

As stated in clause 8.2 of [4], network providers in 2030 will need to be able to rapidly introduce new network services, or network services with properties that can be rapidly adapted to new contexts, deployments, and application needs. Likewise, the business landscape may require users of network services to be able to rapidly adapt services to their needs.0987654er This will require advances in network programmability.

Software defined networking and network function virtualization (NFV) open the possibility of accelerating development lifecycles and enabling network providers to develop new networking features through control and data plane separation. However, modification of controller software is still complex and can only be conducted by network providers. Programmable packet processing technology such as P4 facilitates the rapid introduction of new protocols in support of new services but suffers from many limitations, such as the ability to support variable and dynamic structures.

It should be noted that softwarized networks are built on relatively stable (and slowly evolving) underlying physical commodity hardware network infrastructure. This is insufficient to deliver the network services described in this technical report, which require hardware advances at many levels to provide programmable flow and QoS behaviour at line rate, affecting everything from queuing and scheduling to packet processing pipelines.

Segment routing [19] is being evolved for similar purposes as well. However, there are severe limitations due to the constraint in footprint of any logic or function invocations that can be carried.

6.3.2 Gaps

G.CAP.7. **Rapid customization of network services**

Capabilities are required that allow for rapid customization of networking services for specific needs or adaptation of network behaviour to unique deployments. Specifically, this concerns capabilities which are in reach for network provider customers such as IT organizations. Rapid customization may be seen analogous to a serverless lambda programming paradigm; which when coupled with new network service interfaces can perform rapid customization by embedding application specific SLO processing capability in the networks. Application level over-the-top (OTT) overlays cannot adapt to changing network behaviour fast enough.

In addition, these capabilities need to be addressed in a hardware-friendly manner without prohibitive performance penalties. They also need to be provided in a manner that is secure and does not enable novel attack vectors, for example, those that cannot be used to compromise network provider infrastructure or traffic by other users.

6.4 Manageability

As described in clause 8.3 of [4], support for Network 2030 services will require advances in manageability to be able to successfully provide and operate such services. Such advances will need to proceed in lock step with advances in the services themselves. Gaps stem largely from the fact that manageability today is geared towards best-effort services and itself is "best-effort", relying for example on sampling as a way to scale. However, best-effort manageability is not sufficient for high-precision services. A detailed analysis of those gaps is provided in [7]; the remainder of this clause borrows from that analysis.

6.4.1 Assurance of high-precision services

6.4.1.1 Baseline

The ability to assure high-precision service levels depends on the ability to measure those service levels with high precision and accurately. In addition, the ability to identify and eliminate potential sources of service level degradations and fluctuations will become of increasing importance. This requires instrumentation, telemetry generation, and tracing capabilities which are lacking today despite recent advances.

Measurements involve two aspects. The first aspect concerns the precision and accuracy of the measurements themselves, which need to be conducted with precision in the range order of 10 microseconds for end-to-end services. The second aspect concerns coverage of those measurements which need to be complete enough to detect any violations of service level

objectives. For example, in extreme cases, it should be possible to detect and analyse violations of service levels that may occur in only one of 10^{12} packets.

Today's measurements commonly rely on active measurements, in which test probes generate synthetic test traffic. One problem concerns the overhead that is associated with synthetic test traffic, which consumes considerable resources in sending and receiving probes as well as traffic reflectors, in addition to consuming considerable network bandwidth. For this reason, measurements can only cover a sample of the network at any one point in time. Given the requirements for high precision, the possibility of misses due to statistical sampling may no longer be acceptable going forward. Instead, full coverage needs to be achieved without compromising measurement accuracy.

Passive measurements that rely on observations of production traffic, or hybrid schemes (in which production traffic is augmented with metadata used for measurement purposes), provide an obvious alternative. However, regulatory requirements may stand in the way of solutions, as observation of production traffic is required, which causes privacy concerns. Likewise, achieving the required measurement accuracies while avoiding performance hits on that production traffic, can become challenging.

With regards to instrumentation and generation of telemetry data, important advances have been made in recent years. For example, YANG-Push allows subscription to continuous streams of network device data and statistics. Distributed network analytics allows dynamically adjusted data to be generated at the source as needed, obtaining more meaningful and actionable service assurance data. However, again there are limits to the frequency with which data snapshots can be obtained using existing instrumentation technology. Likewise, the ability to collect comprehensive data at a per-packet level using in-situ OAM, while highly useful, is still limited. Issues include impact on performance to retrieve the data, as well as the sheer potential volume of data (one data record for each packet and hop) coupled with severe limitations in the amount of data that can be piggybacked due to MTU considerations.

A greater emphasis will be placed on the ability to monitor, observe and validate compliance of actual with expected network behaviour than is the case today. Data to be generated from the network should be more insightful and actionable. This will require additional abilities to process data "on-device".

6.4.1.2 Gaps

G.MGMT.1. **High-precision measurements**

There is gap in measurement techniques that are capable of measuring end-to-end service levels that combine all of the following. Accuracy: The combination of high measurement accuracy (precision on the order of 10 microseconds end-to-end) with acceptable measurement overhead; Coverage: lack of measurement techniques that allow for complete coverage not only of selected service instances but across the board, specifically for mission-critical, high-precision services at scale, without having to resort to sampling; and, Privacy-preservation: no reliance on packet snooping that could expose personally identifiable information.

G.MGMT.2. **Comprehensive device telemetry at scale**

There is a requirement for techniques that allow the generation of actionable telemetry on individual devices, with the telemetry being fine-grained enough to allow for analysis of individual packet behaviour traversing the device and to able to operate at very high frequency (i.e., able to provide updates at ms or µsec frequency). Such techniques are needed to assure high-precision services.

G.MGMT.3. **Comprehensive end-to-end packet telemetry at scale**

Techniques that allow the collection of telemetry data, end-to-end across network devices, for given packets, flows, and service instances, in a way that generated telemetry does not overwhelm or

otherwise negatively impact production traffic, and that does allow to scale, i.e., to cover every high-precision flow and flow burst of interest.

G.MGMT.4. **Service level validation**

There is a gap in technology that allows continuous validation of whether service levels being delivered adhere to agreed-upon objectives. This will be required with the transition towards high-precision services that are delivered at a premium, both to assure and to properly account for those services. With best-effort services in the past, that was much less of an issue.

Another set of concerns for both users and providers of network services concerns the possibility of data and privacy leakage. As Network 2030 applications are becoming increasingly mission-critical, and as the sensitivity to, and impact from, privacy attacks is increasing, the importance of management functionality to address this is growing.

G.MGMT.5. **Assessment of data or privacy leakage**

Network providers need to assess whether leakage of data or privacy from communication flows of users is occurring. Specifically, there is a lack of technology that provides evidence for the possibility of presence or absence of such leakage.

6.4.2 Fulfilment and "operation-at-scale"

6.4.2.1 Baseline

Another challenge involves enabling operators and users to manage Network 2030 at scale. The growing need for shorter control loops means that management services will increasingly need to migrate closer to the edge of the network and indeed into devices themselves.

However, despite all those advances, networks will not become clairvoyant and will still need to be given guidance for certain tasks and require some degree of human interaction. For this reason, advances in abstractions will be required to facilitate the ways in which operators can interact with networks. These abstractions are needed for productivity reasons (operate at greater scale) and to constrain complexity (greater heterogeneity, growing number of interdependencies which are becoming less understood, etc.). It is here where the biggest gaps exist today. Technologies such as intent-based networking, which will allow networks to be managed by defining outcomes rather than prescribing rules or procedures, are expected to provide significant contributions here but are still in their infancy.

6.4.2.2 Gaps

G.MGMT.6. **Human/machine interfaces for intent**

There is a technology gap with regard to human/machine interfaces that allow users and networks to iteratively infer and refine intent, based on the definition of desired outcomes as opposed to requiring users to define specific courses of actions to take.

G.MGMT.7. **Automated planning of explainable courses of actions for outcomes**

Current systems are good at automating tasks and using predefined algorithms, but there is a lack of systems that are able to automatically define and continuously refine plans of actions that generate desired outcomes, and are able to explain their actions.

G.MGMT.8. **Decentralized and distributed architectures for low-latency management at scale**

Management architectures that are able to scale without boundaries and that are able to support management and control loops at extremely short time scales – milliseconds or microseconds – as will be required for "self-driving networks". Specifically, there is a gap in systems that support greater management functionality in a distributed or decentralized manner across the network, as

opposed to relying solely on centralized management systems and controllers as is predominantly the case today.

6.5 Security

6.5.1 Baseline

Security is an ongoing concern for Network 2030 as well as todays networks. Many gaps exist even today. Most notable is perhaps the absence of authentication, which facilitates spoofing and a myriad of attack vectors such as phishing and distributed denial-of-service (DDoS) amplification attacks. These gaps continue to carry forward.

Network 2030 services will need to be secure. Network services such as coordinated communication services or qualitative communication services defined in clause 5, can be originated at a network ingress point and consumed from an end host or network egress point. Additional security mechanisms are needed beyond those that are provided by traditional transport mode IP security.

6.5.2 Gaps

G.SEC.1. **Authorization of packets**

Networks need to verify the degree of integrity and whether packets are authorized to enter into the network. However, when packets are emitted from the host for these new communication services, the network portion of the packet (e.g., an extension header or an overlay header) should not be encrypted unless network nodes can still interpret the header and provide the desired service. Lack of encryption and integrity validation, of course, would at the same time increase the threat surface and open up the possibility for attacks. Mechanisms for authorization and integrity protection must be developed to meet the line rate performance as services delivered can be time sensitive. At the same time, the size of packets should not be significantly increased to avoid negative impact on utilization and overhead tax. This limits the options for additional security collateral that can be included with packets.

G.SEC.2. **Authentication of packet headers and prevention of spoofing**

Networks need to provide mechanisms that rule out the possibility of spoofing, for example, impersonating a different / unsuspecting source as a sender. This means that the information in packet headers that identifies a sender or source (e.g., a source address) needs to be authenticated. This is a serious gap even today, whose severity will be compounded for Network 2030.

Note that this does not preclude the possibility for communication that is anonymous. However, anonymity is to be differentiated from impersonation. Senders may be anonymous by being transparent about the fact that they are anonymous, i.e., not pretending to be somebody else.

G.SEC.3. **Homomorphic encryption**

Homomorphic forms of encryption may need to be devised that allow network operations to be performed in a privacy-preserving manner on encrypted packet headers and tunnelled packets without exposing any of their contents. Homomorphic encryption of packet headers ensures that applying an operation to the encrypted packet header will result in the same result as if the packet header had been decrypted, the operation performed, and the packet subsequently be re-encrypted. This will help also keeping all communication that occurs in a network private, while still allowing the network to function properly.

(NOTE – This is also a requirement for privacy).

G.SEC.4. **Network 2030 services security: Coordinated services**

By attacking a single member flow, the co-flow as a whole as well as other member flows could be compromised. For example, an attack that introduces additional latency on a member flows might also slow other flows depending on inter-dependencies.

G.SEC.5. **Network 2030 services security: Qualitative communications**

With qualitative communication services, it is no longer enough to secure packets and ensure their integrity as a whole. Because chunks, i.e., certain portions of the payload, might be legitimately dropped, packets and payload also need to be secured at the individual chunk level. Such mechanisms are not available today.

6.6 Resilience

Resilience is the ability of the network to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation.

At the network (topology) level, resilience amounts to preserving loss, jitter, and latency as successfully as possible for a given service – all these quality of service (QoS) metrics can be compromised if failures/attacks occur and there is a lack of resilience mechanisms to remediate/mitigate them [18].

6.6.1 Baseline

Today network providers support resiliency in their infrastructure under four main elements: by providing path geographical diversity, redundancy through back-up for equipment, protection to minimize threats that can affect the network, and restorative for rapid recovery from loss, outage or congestion. There are different approaches to these methods and usually it is impossible to infer if the overall network-system is resilient. To provide redundancy and diversity of logical and physical entities such as network paths, functional entities, ports, routers, and router line cards. The routing protocols are used to provide fast re-convergence through mechanisms such as fast reroute (FRR) in MPLS and IP to support link and node protection. Faster convergence is achieved using pre-computed loop free alternatives (LFA) repair paths. Among the other techniques is the use of packet replication, network coding, and error correction to overcome packet loss.

6.6.2 Gaps

Resilience needs to be a system level capability which in turn is a combination of several other capabilities discussed. Resilience takes on additional importance for Network 2030 services, because in many cases these services are used for mission-critical applications and require high-precision, moving beyond "best effort" that was acceptable for many applications in the past.

G.RES.1. **SLO-aware resource protection**

Network 2030 services are characterized by the need for high precision timing (e.g., in-time and on-time services) and synchronization between large numbers of flows (coordinated services). Any network degradation puts these services in jeopardy and makes the applications that rely on them infeasible. The ultra-low-latency requirements, and the huge increase of bandwidth demands of Network 2030 services such as holographic type communication services, make an unrecovered failure a significant loss for network operators. The switchover from the primary entity to the backup entity must be very fast in the order of micro-seconds or even less.

G.RES.2. **SLO-aware path diversity and protection**

When traffic is rerouted from the primary path to a new path, the new path should provide the same network high precision communications services that are available in the primary path.

G.RES.3. **Resilience intent and levels in SLAs**

Appropriate resilience objectives that indicate the importance of the service in terms of expected availability and reliability, along with articulation of acceptable trade-offs. This statement of intent will be mapped into the additional resilience measures to be taken to avoid violating the SLAs.

G.RES.4. **Resilience at system level**

Resilience capability resides in network design or in the architecture domain. Here resiliency is mainly concerned with the requirements of the new services in Network 2030. Such services requiring resiliency need to know what kind, or how much, they can depend upon an underlying infrastructure. This requires a network operator to have runtime knowledge of the network state at any instant which in turn relies on continuous collection of metrics. Therefore, measurability, security, and other capabilities covered are necessary to quantify resilience through a combination of all its aspects.

G.RES.5. **Resilience control knobs**

Higher resilience can be achieved by providing spare capacity along backup paths, which results in lower utilization. Providing mechanisms that allow users and providers of network services to manage those trade-offs.

The network resiliency itself is achieved through a multi-layer approach. The study for this clause is not complete.

6.7 Loss-lessness

Loss-lessness capability offers near zero packet loss as an objective for mission-critical applications that cannot afford to rely on retransmission and reliability schemes. While no service will be able to guarantee zero loss due to the possibility of some catastrophic event, loss because of single equipment or link failures or due to congestion should be ruled out.

6.7.1 Baseline

A major source of packet loss is tied to the classical problem of congestion and limitations in network resources (bandwidth, buffer spaces) because of competing flows. One way to avoid loss is to avoid oversubscription of resources to ensure that congestion cannot occur. However, in general this leads to poor network utilization and physical link failures can still occur. In order to protect against the possibility of link or equipment failures, in IEEE P802.1CB *Frame Replication and Elimination for Reliability,* network traffic is sent redundantly over multiple paths.

6.7.1 Gaps

The challenge concerns how to achieve loss-lessness while keeping costs acceptable.

G.LLS.1. **Cross-layer support**

Each layer in the network stack provides its own mechanisms to prevent packet loss. Therefore, a higher layer would also need to request loss-lessness as a service capability from the layers below. The assumption is that higher layers makes loss-lessness criteria known to lower data and physical link layers which then use their capabilities to deliver packets.

G.LLS.2. **ML techniques predicting congestion**

Possible mitigation techniques include machine learning and AI techniques, to recognize the possibility of resource contention early and to dynamically adjust packet forwarding in such a way that congestion and thus loss are avoided even at high utilization levels.

G.LLS.3. **Application defined lossless-ness criteria**

Applications should be able to specify criteria about the data loss. It contains guarantees of packet loss rates. This is primarily useful for media applications for their consistent quality of experience. Further utilizing finer granularity of service level objectives, applications should be able to describe

on a per packet basis which ones are critical and must not be dropped. Extending this concept further through qualitative communications, the loss-lessness criteria may be associated with a part of the data stream or a part of the packet.

6.8 Privacy

This clause refers to the privacy of data in transit and corresponding control information in the network.

6.8.1 Baseline

In general, routing services rely on information in data headers of packets to provide their service. However, the combination of data analytics, with the PII such as addressing and third party sharing of information create an opportunity to track the user and observe patterns even in the presence of encrypted data. This is very much an open problem today.

A possible option is to provide privacy through anonymity, for example through use of onion routing (TOR) as used on the darknet. However, such solutions trade anonymity against security and do not support service level guarantees, making them generally non-acceptable. The solution to greater privacy and protection of the PII will need a technical solution at the network level as well as regulatory solutions.

6.8.2 Gaps

In addition to ongoing efforts on privacy, of particular relevance to Network 2030 are:

G.PRIV.1 **Opacity of user data**

Network 2030 services must not rely on the user data to provide the service but rather on specific service-visible data in the packet. For example, this information may be the service level parameters for the data in the packet. These parameters are distinct from the user data which need to be opaque.

G.PRIV.2 **Protecting service level parameters**

Service level parameters can be interpreted by the network devices. Today most information in headers is clear and therefore observable by eavesdroppers. The mechanisms behind which in-network parameters are protected, kept to a minimum or encrypted when possible, require further research.

G.PRIV.3 **Secured storage**

A lot of Network 2030 services and use cases such as NCC may require data in-transit to be temporarily at rest on the router. The storage of those packets should be secured in such a way that it is not duplicated, stored, or undergoing deep inspection unless authorized. Mechanisms to detect any tampering or exceptions are topics for further research.

G.PRIV.4 **Flow anonymization**

Data should be obfuscated but the flow of information should be randomized in a dynamic manner so that it is difficult through traffic analysis to deduce patterns and identify the type of traffic. Services such as qualitative analysis may provide more fluidity in the traffic patterns for hard correlation.

6.9 Validation of delivered services

Many Network 2030 services place very high demands on the network in terms of required service levels, demanding guarantees instead of being accepting of "best effort". Guarantees demand their price, making it increasingly important to be able validate that the promised service levels were delivered.

6.9.1 Baseline

Today's accounting technology largely relies on interface statistics and flow records. Those statistics and records may not be entirely accurate. For example, in many cases their generation involves sampling and is thus subject to sampling inaccuracies. In addition, this data largely accounts for volume but not so much for actual service levels (e.g., latencies, let alone coordination across flows) that are delivered.

Service level measurements can be used to complement other statistics but rely largely on active measurement techniques that also have limitations related to sampling. In addition, it comes with significant overhead, including the consumption of network bandwidth as well as additional processing on edge nodes. Techniques that rely on passive measurements are unfeasible in many network deployments and hampered by encryption, as well as issues relating to privacy, the concerns for which are expected to increase further.

6.9.2 Gaps

Validation requires advances in accounting technology.

G.VAL.1 **Accuracy**

Measurements of service levels will need to be accurate enough to account for high-precision service performance targets and be provided at scale for all high-precision communication instances. Statistical methods that rely on sampling as a way to scale, and that may miss violations of service level objectives, will no longer be acceptable.

G.VAL.2 **Proof of service delivery**

Proof of service delivery (including proof of service level delivery) may need to be provided to account and charge for network services. This is particularly the case as network services move from best effort basis to a guaranteed basis and are used for mission-critical applications. Guarantees should be expected to have their price, and best effort accounting may no longer be sufficient for 2030 networks.

G.VAL.3 **Accounting advances**

Advances in accounting methods are needed in order to enable new incentive-based schemes to deliver services. For example, using prepay models, applications would no longer be able to just demand a network service with a certain network service level and rely on the goodwill of the network provider to provide it, but give network providers an incentive to deliver on them. Conversely, network providers will be able to allocate their resources more effectively than today in ways that best support economic goals. This can enable new business models and communication service supply chains that in turn foster further innovation for Network 2030.

G.VAL.4 **Accounting for new network services**

Existing accounting methods that operate at the level of packets and flows are not able to deal with novel network services. Qualitative communications introduce a new level of communication units referred to as "chunks"; accounting methods need to evolve to be able to properly account for chunks. Likewise, coordinated communications will require the development of techniques that account not just for individual packets and flows being delivered, but for their coordination.

7 Analysis of network infrastructure and operations

This clause deals with gaps associated with the support of Network 2030 considering deployment, infrastructure and operational requirements.

7.1 Compute in networks

For the applications requiring low-latency and energy consumption reduction, it is necessary to process at least some of the data at the edge rather than sending them to the remote data centres.

7.1.1 Baseline

Development of cloud computing concepts allows end devices to exploit computation and storage resources hosted by powerful farms of servers. At the same time, high latency gets introduced due to the fact that the users are far away from the servers in terms of network topology. Real-time requirements or interactive behaviour for many emerging applications will not accept latency performance degradation.

Edge compute networks provide the support for compute in networks generically and address the issue of long latency by offering services to the proximity of the end-users. The edge compute proposals include Cloudlet [8], Fog computing [9], Mobile Edge computing (MEC) [10][14] (Multi-access Edge computing[11]) and Mist computing[12].

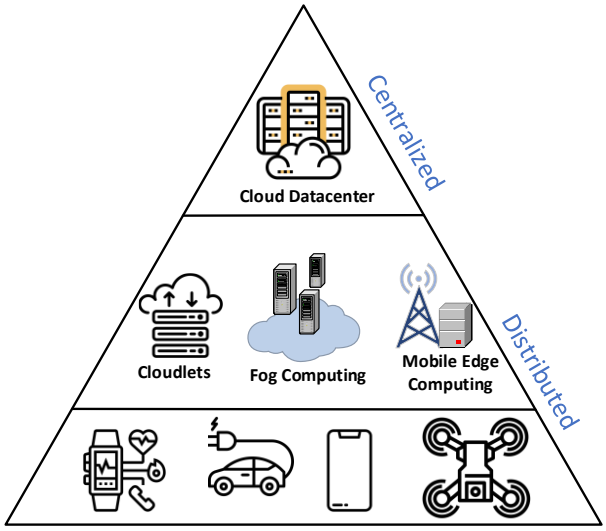

Figure 3 – Multiple-tier compute in the network

Traditional cloud services assume client-server style interactions which only have two tiers. However, with the introduction of distributed edge computing and end devices, it entails at minimum a three-tier architecture with its own set of coordination and orchestration requirements as shown in the Figure 3. Together they can provide heterogeneous computing resources for different applications scenarios.

The computation resource allocation and scheduling are complex processes influenced by different factors, such as users' preferences, radio and backhaul connectivity, mobile device's CPU frequency and power residual, servers' computation capability and availability. Current research [15], [16] and [17] mostly considers one or two factors and develops the algorithms with centralized control, neglecting the fact that the servers in the intermediate tier in Figure 3 are deployed in a distributed fashion. Moreover, the characteristics of the underlying network such as topology and bandwidth resource should also be considered.

Networks operations would need to provide mechanisms to access edge compute nodes that best serve the application resource requirements.

7.1.2 Gaps

For compute in networks, there are gaps in the control and management of computing resources, the coordination between network and computing resources, and the reliability of computing resources.

G.NCC.1. **Computing-aware routing**

To meet the distributed computing demands and resource availability in the network, support for computing-aware routing will be needed. It requires run-time considerations for attributes such as geo-location (physical proximity) and the status of computing resources when routing the traffic flows and computing tasks to the optimal computing site.

G.NCC.2. **Coordination among different computing resources**

There will be different scales of computing resources in the future networks, from large-scale data centre, to small-scale edge data centre, as well as intelligent devices with limited computing resources, each has various kinds of characteristics that need to be taken into consideration. Emerging applications (for example, AR rendering, autonomous driving) are characterized by high mobility, high computing complexity, and strict time constraints. To guarantee differentiated service experience with much higher granularity than in current networks, coordination within the edge tier or between cloud and edge tiers are required.

G.NCC.3. **Reliability of computing resources**

Future networks should establish a reliable authentication mechanism to make sure that the computing resources assigned to provide services are secure and reliable, to protect security and privacy of the user accessing the service.

G.NCC.4. **Service requirements awareness**

To meet the service requirements, including the network requirements and computing requirements, it is not enough to be aware of the status of computing resources in the network, the awareness of service requirements is essential. By combining the exact service requirements, including latency, jitter and security, etc., and status of network and computing resources, the network can define the optimal scheduling policy guaranteeing the service quality.

G.NCC.5. **Distributed and unified framework**

There is a lack of a unified and distributed framework that realizes the user-oriented service in the network. Such a framework needs to be able to incorporate the users' preferences on the characteristics of various compute resources along with the properties of the network. A collaborative compute scheduling framework is needed among the edge-compute networks through which they offer and request resources with the given set of service attributes. Vertical service-specific management is necessary that takes advantage of ubiquitous computing resources at scale enabling autonomic management, without manual intervention.

7.2 Intelligent operation networks

Intelligent operation networks are a fully automated and intelligent closed-loop control framework to monitor network health while supporting flexible and complex network functions.

7.2.1 Baseline

Clause 6.4.2 covered state-of-the-art management aspects of the network operations. Regarding intelligence in networks, intent based networking (IBN) [20] is being defined to provide high level operational objectives and outcomes in a declarative manner. The rendering of an intent depends on management APIs and service specific data models.

7.2.2 Gaps

For intelligent operations the key performance indicator is "network health". To derive this KPI, use of multiple sensors in the network is required to continuously measure numerous parameters. The IBN, ZSM [21] frameworks need to evolve for Network 2030 as follows:

G.ION.1. **Massive data computing, storage, collection and analytics**

There is a need for intelligent and efficient methods to store and compute the massive data required by ION to truly cognize the network and train AI models. The gap needs to provide initial (or starter) AI models for network availability, utilization, throughput, congestion states (links, routers/switches) and transit path of a traffic, etc. These models can then be modified, learnt and adapted to a specific network operator's environment.

The scale and rate at which telemetry data needs to be collected from the network for high-precision services will be much larger. Therefore, techniques in which distributed ML-processing is utilized are necessary.

G.ION.2. **Intelligent and low-latency reporting**

Network 2030 should further extend data reporting elements. Telemetry data should provide anomaly-reports with low latency for fast response. Instead of reporting high-volume of raw statistics, the network elements may report intelligent and accurate data to the control-centre. Reported data should be structured and re-labelled to facilitate analysis by the control-centre.

G.ION.3. **Autonomy**

In order to complete the closed-loop of intelligent control, all the elements in the system need intelligence to process and analyse the data collected. In order to recover fast and enhance reliability, these elements should be able to take optimal measures in time by means of artificial intelligence and big data analysis.

7.3 Support for ManyNets

ManyNets is an existing phenomenon described as existence of several very large-scale and global-reach networks. The purpose is to identify inter-connectivity challenges across such independent large-scale networks.

7.3.1 Baseline

While we have witnessed rapid technology development in each specific type of network such as fixed Internet, 5G, satellite networks, etc., their evolution into a holistic network ecosystem with harmonized coexistence is not yet clear. This is particularly challenging when we consider that different standardization bodies are in charge of networking technologies in specific network scenarios (e.g., IETF on Internet protocols and 3GPP on 5G). On the technical side, network-specific mechanisms and protocols have been proposed and standardised but until now there has been little progress on how they interoperate seamlessly across boundaries between heterogeneous networks. For instance, a wide variety of dedicated routing algorithms have been proposed for low-earth orbit (LEO) satellite networks, making sure the paths through chains of LEO satellites are optimally maintained over dynamic mega-constellation behaviours. To interwork with the network infrastructure on the ground, the common practice is to deploy protocol translations or packet encapsulations at the network boundaries, making the end-to-end data path less natural. Similar stories take place between cellular and fixed networks where packets are delivered through GPRS tunnelling protocol (GTP) tunnels before being injected into the public Internet. We believe that such fragmented network technologies and protocols substantially hinder the future evaluation of unified ManyNets. The key issue is that, based on environment-specific protocols it is difficult to provision end-to-end QoS for end users across network boundaries when ManyNets are leveraged for delivering the traffic. A similar case can also be made for end-to-end security and resource management across different network environments.

7.3.2 Gaps

G.MNS1. **Flexible routing architecture**

The convergence of ManyNets will eventually witness unified addressing and routing architecture. In particular, most of the proposed LEO satellite networks in the literature are based on dedicated non-IP based mechanisms, typically through tunnels. In this case, at the boundary of terrestrial and space networks, protocol translation and/or packet encapsulation operations need to be in place in order for the user data to be able to traverse different networks with heterogeneous protocols.

G.MNS2. **Integration of transit and access functionalities**

Technically, LEO satellite networks can be used for both direct user device access and transit services. On the other hand, most of the currently proposed network architectures / protocols are only based on one of the service scenarios. In this context, the realisation of a common network framework that enables simultaneous support of access and transit network functionalities will become essential to seal the technology gap. It should not be required to develop two separate set of protocols or solutions for access and transit roles. The main challenge here is the overall system architecture where network functions dedicated to different functionalities can be systematically and seamlessly integrated to support both service scenarios.

G.MNS3. **Quality of service support**

While the integration of heterogeneous networks is expected to increase the overall network capacity for handling Internet traffic, the provision of end-to-end QoS support, typically across network boundaries, is yet to be investigated. Concerning the LEO satellite network, it is essential to consider the physical capacity of the inter-satellite links (ISLs) as well as the downlink and uplink characteristics between the satellites and the terrestrial ground stations. From an end-to-end point of view, to the question on how to compute QoS-constraint paths within LEO satellite networks (mainly satellite mega-constellation behaviours) is a key challenge. Additionally, how we cross the boundary of terrestrial and space networks represents another key challenge in the context of QoS support. This also involves the up-to-date awareness of the dynamic traffic load conditions in different networks which requires sophisticated distributed network monitoring, admission control and dynamic traffic steering against network uncertainties.

G.MNS4. **Resource management**

The traditional view on network resource management has mainly focused on communication network resources, in particular upon bandwidth resources. While this will remain a major concern for LEO satellite networks in space, it is also envisaged that, with the future possibility of on-board data processing and storage capabilities in LEO satellite networks, edge caching and computing could become a new feature. In this case, the scope of resource management will need to be expanded to cover new IT capabilities from space in conjunction with traditional communication resources.

7.4 Artificial intelligence aware networking

AI aware networking refers to connectivity and placement aspects of AI components (data, models, knowledge) for the deployment of intelligent services along the cloud-to-things continuum.

7.4.1 Baseline

Connecting a chain of computing resources for executing generic computing tasks at the edge has been a subject of recent research efforts [22]. However, the orchestration of AI components encompassing the placement, the connectivity and chaining cannot be tackled in the same way.

A proper understanding of AI peculiarities is required when designing networking procedures, which need to be conceived to be natively AI-aware.

The accuracy and privacy demands of AI applications should be considered when deciding where to place the AI training and inference procedures, in addition to the common KPIs, such as high energy efficiency, low latency, that are already driving existing edge computing solutions. On one hand, pushing AI models close to data is crucial to ensure privacy preservation. On the other hand, heterogeneous devices may provide inference results with different accuracy levels, e.g., according to their capabilities. Moreover, network procedures should account for the dynamics of the considered AI algorithms, e.g., the number and type of contributing nodes for the sake of model building [24].

The pervasive distribution of AI capabilities to end-devices, network nodes, edge/cloud facilities imposes new challenges to the design of Network 2030.

7.4.2 Gaps

G.AIN1. **Extended discovery and addressing mechanisms**

In a pervasive AI deployment, every entity (i.e., IoT device, edge network node, data centre) can contribute to the AI workflow, by providing raw data, inference capabilities or model updates. Conventional host-based addressing and reachability do not match the targeted scenarios. Applications should be able to discover and connect with specific AI components, especially DL models and inference capabilities provided by heterogeneous and distributed devices along the cloud-to-things continuum. According to the targeted applications, such components can be chained for the sake of service composition. Hence, the network should identify and discover specific AI components through their unique capabilities or identities.

G.AIN2. **Joint network, intelligence and computing orchestration**

Efforts are underway, but still in their infancy, to synergize allocation of computing, caching and communication (3C) resources [23]. However, the orchestration of AI components goes well beyond existing joint 3C solutions. Algorithms should be conceived which, beside computing resource availability, network topology and link conditions, specifically account for DL model capabilities to satisfy application demands, e.g., in terms of privacy (whenever personal/sensitive data are given as inputs to training models) and accuracy. Moreover, the popularity of requests for some specific models/inference results, typically exhibiting a spatiotemporal locality, should be accounted to drive caching decisions.

Personal/sensitive data may be required to perform model training procedures feeding smart applications. Among DL solutions, federated learning enables local training on edge/mobile devices without personal data exchange between the server and clients, thereby protecting clients' data from being eavesdropped by hidden adversaries. However, private information is still susceptible to leakage by analysing the differences between the parameters trained and uploaded by the client. To mitigate such leaks, for example, differential privacy with artificial noise added to the parameters notified by the clients can be applied. Hence, the suitability of specific privacy-preserving AI algorithms as well as their actual robustness to privacy attacks should be evaluated whenever needed.

G.AIN3. **Group-based communications**

Point-to-multipoint communications are expected to be common whenever edge/mobile devices, equipped with AI capabilities, need to be simultaneously queried. This could be when required to perform some training tasks (e.g., in the case of federated learning where devices are requested to locally perform model training) or be the simultaneous recipients of updated AI models (e.g., multiple surveillance cameras in a smart city, client devices in the case of federated learning). Traditional communication primitives, both multicast and broadcast, either relying on network-layer procedures or server-assisted group-based communications, barely match the aforementioned scenarios. Novel network primitives are required which allow to recognize the most suitable entities to be reached and efficiently forward data to them accordingly.

7.5 Support of SIoT-enabled logistics

7.5.1 Baseline

The exponential growth of e-commerce along with the novel strict regulations regarding the circulation of vehicles in urban areas and the need for cost effective operations are raising unprecedented challenges on the logistic sector. Such challenges cannot be dealt with without the full support of ICT solutions which are able to gather massive amounts of relevant data from a multitude of heterogeneous information sources managed by a diverse set of organizations. To this purpose, recent logistics solutions, such as COG-LO (http://www.cog-lo.eu), exploit the social Internet of things (SIoT) concept which envisions smart sources of information (i.e., the *things*) to establish social-like relationships with each other almost autonomously. Such contexts introduce several new pressing demands which cannot be satisfied by current networking technologies. In fact, while the upcoming 5G networks are expected to be able to cope with paramount quantities of data which will need to be exchanged and with the low-latency requirements, there are aspects of SIoT-enabled logistics which require a new generation of networking solutions.

7.5.2 Gaps

G.SIOT.1. **Things virtualization**

Most Internet of things (IoT) approaches (including the SIoT) envision the use of digital counterparts to augment the capabilities of physical objects, which are often called *virtual entities*, each one representing and acting on behalf of an object in the digital realm. Current solutions implement the virtual entities as agents that run in application level platforms, each one with limited capabilities to efficiently interact with the other platforms so failing to fulfil the core IoT promise to have any object capable to interoperate with all the others. New networking solutions are required that define virtual entities as processes hosted inside the network infrastructure by offering the needed environment along with the interfaces needed for interacting with them. This requires the design of appropriate reliable, trustworthy, scalable protocols along with the deployment of the required computing and communication resources. This would allow entities to interact with other entities, network services, and application layer processes.

G.SIOT.2. **Support of application-specific communication primitives**

While the network has changed radically in recent years from many points of view, at the very end the communication models it supports are all based on unicast, multicast, and broadcast communication primitives. There are cases, instead, in which new communication primitives are needed. This is the case of *sociocast*, which is a social-driven networking primitive that enables trusted group-oriented communications. It allows transmitting devices to define the group of recipients (and recipients to define the list of devices that can send data to them) on the basis of their social distance, which is computed from a social network generated and maintained by the network [25]. This primitive relies on the SIoT paradigm. Supporting new application-specific communication primitives requires the definition and support of abstractions of communication primitives and their support through programmable control and the data planes.

G.SIOT.3. **Mobility of things and virtual entities**

While mobility support has been the focus of large research and development efforts in the context of cellular networks for three decades, this scenario requires the design of totally new solutions which take its unique features into account. In fact, the aspects to be considered simultaneously are the need to keep the characteristics of thing and its digital counterpart as close as possible, the constraints in terms of computing and communications arising from the limited resources available in the things, and the existence of social-like links between things that can be used to predict the communication patterns. Indeed, social-like links that arise from the SIoT functionalities implemented at the network layer [24] can be successfully used to predict the future patterns of interaction among the devices. Data exchange is likely to occur among devices tied by social

relationships (e.g., sharing of traffic information among vehicles usually traveling along the same roads). In addition, the stronger the social links the higher the likelihood of frequent interaction among devices.

G.SIOT.4. **Differentiated security and trust for resource constrained devices**

Current networking solutions do not exploit the social-like relationships between objects established in the social Internet of things. This, instead, could be extremely beneficial because security procedures are extremely costly in terms of energy, computing, and communication resources. It is therefore desirable to envision a new generation of networking solutions that adapt the effectiveness of security procedures (and therefore, their costs) to the context in terms of the positions in the social graph of the involved parties.

8 Evaluation of key findings

This clause provides overall analysis on the basis of gaps.

8.1 Use cases to services and capabilities relevance

The work on use case reports and Network 2030 services was carried out in parallel and this clause serves to provide a certain degree of correlation between them.

The study of use cases can map to new services and capabilities presented in the Table II.1 in Appendix II. We utilize a simple method, assigning a relevance score of high, medium and low. The row headings refer to use cases and the acronyms in the columns denote services and capabilities. A simple observation can be made in terms of the level of dependency a particular use case has on a services or capability.

Different clusters of use cases have overlapping requirements, this is shown in Figure 4 showing dependency between services and the use cases. It is separated in to 3 parts: a) service provider, b) industry and c) common for reasons of clarity. The size of the nodes is based on their degree and weights. The bigger the use case node size (top), the higher the dependency on services (bottom). Similarly, the bigger the node size means a greater number of user cases rely on them. Thus, we observe that HPC, reliability, lossless-ness and tactile network services are crucial for the support of Network 2030.

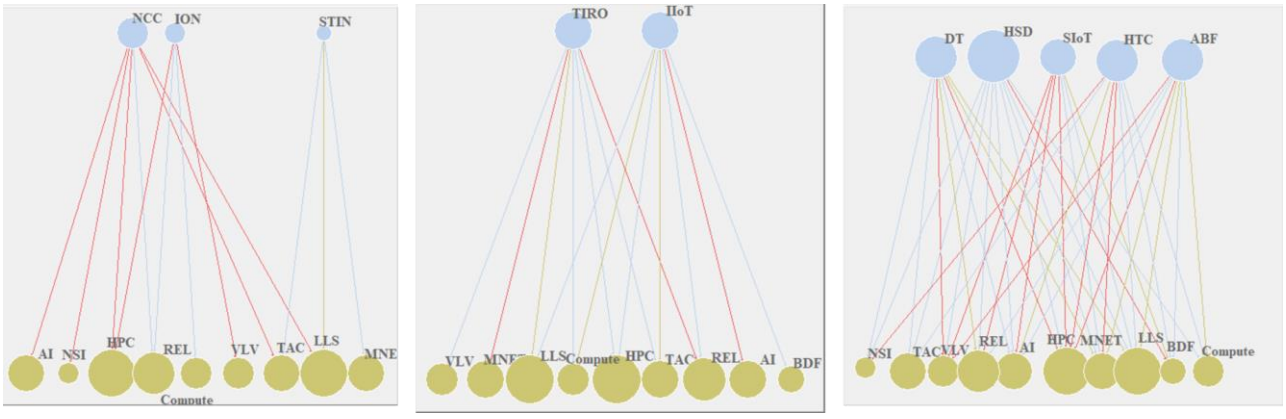

Figure 4 – Use case to network services mapping a) service provider, b) Industry centric, c) common use case

8.2 Factors considered for overall analysis

The Network 2030 effort is not straightforward as it is impossible to focus on one aspect or functionality. Instead it requires a thorough, holistic understanding of all aspects of networking. The gaps or missing functionality were found to be considered along the following factors:

- **Technology gap**: This factor refers to the advancement or improvements necessary in the hardware or software design or paradigms for a particular gap. For example, in some cases technology does not exist or is not available to implement a specific algorithm commercially. In another example, in order to support AI in small devices, challenges relating to energy and processing capabilities need new hardware centric innovations.
- **Algorithm type gap**: This factor refers to the requirement for devising new mechanisms and further study of a solution or formal logical procedures required to support a Network 2030 feature or service. For example, which type(s) of AI models are relevant for future network operations, or would fill the gaps identified in coordinated services, and therefore need to be verified thoroughly following initial development? Questions such as these that require addressing, fall into this category.
- **Architectural gaps**: In the context of Network 2030 work, a separate network architecture study will be conducted. We only observe architectural factors in the scope of gaps discussed in this technical report. Factors that impact at large scale relating to design, deployment, changes to interface between end hosts and network nodes, etc., should be studied under architectural gap. For example, compute in network needs storage as an intermediate node between source and destination and is an architectural gap as current networks were designed for data at rest. Many applications demanding a closer interaction with the network are also an architectural gap (considering how transport mechanisms are an end to end principle) requiring changes in the design and interface to networks.
- **Protocol type gaps**: This factor identifies the aspects where rules for communication need standardization or a common format. Aspects relating to programmability, interfaces between different entities and orchestration are covered by protocol type gaps.

The remainder of this clause is primarily statistical analysis on the list of gaps based on classification criteria (see Appendix II).

It was not possible to classify a gap in to one criterion. Instead each gap is associated with the significance of each of the factor as high, medium, low, or nil and called it weighted gap-factor. Every gap was evaluated for the above 4 factors (based on description of the text) and corresponding weights were attached. When a gap is labelled 'high' for a factor, it means that there is a wider gap along that factor, i.e., more work will be needed to fulfil that gap. Or in other words, solutions will require study, developments or improvements along that factor. When labelled 'low', only a small functionality of that gap depends on that factor.

There are 73 gaps identified in this technical report. The following observations are made:

a) Figure 5 shows all the gaps with at least one 'high' gap-factor on any of the factor – Architectural, Protocol type, Algorithm type, and Technology. The degree of edges into Technology and Algorithm type nodes shows that gaps in Network 2030 services and capabilities is much wider and will wider require more focus on technology and algorithm development.

b) We observed from the total number of labels that the distribution between low, medium, high labels across all 4 factors was found to be 41%, 41% and 16% respectively. Which says that a smaller number of gaps have very "high" dependency on a specific factor. The efforts will need close coordination from all factors. The details are captured in Appendix II.

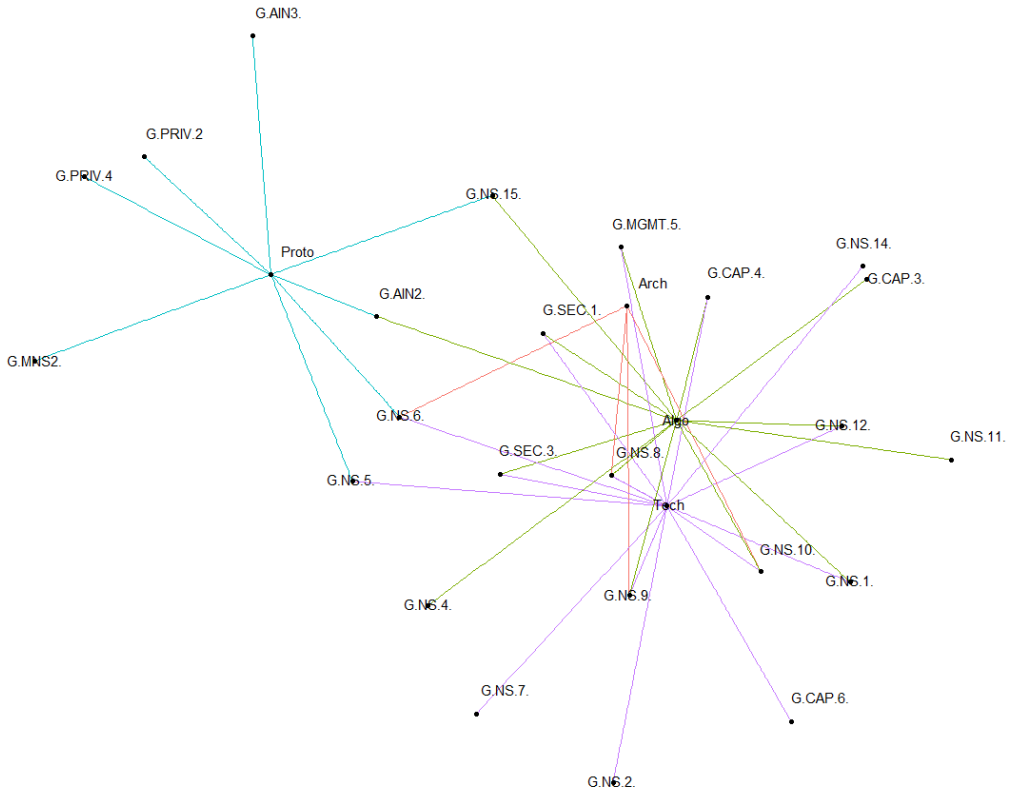

Figure 5 – Relevance-weight=High for architecture, protocol, technology and algorithms vis-à-vis gaps

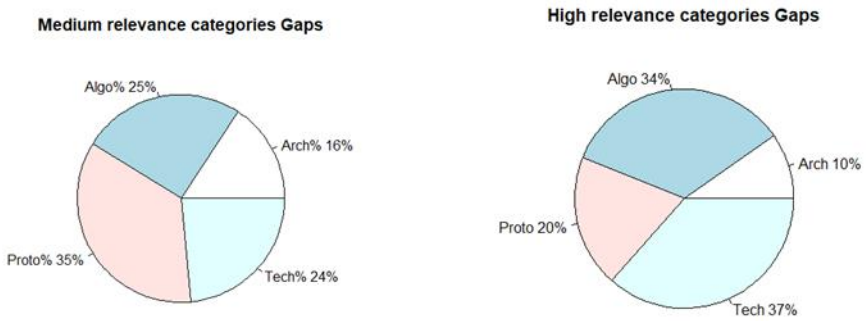

Figure 6 - Distribution of gap-weight-factor a) Average and b) High gap among different factors

c) Another way to observe even distribution of gaps is shown in Figure 6, the first chart shows on an average how the gap distribution across the factors look. On an average, more emphasis is on protocols even though technology and algorithm related gaps are significantly high. In the second chart, the percentage of the widest gaps is shown, and we observe that technology and algorithm become more significant.

d) The significance of compute and AI is observed is several categories. Not just for enhancement of an existing capability but as core functionality to providing network intelligence and convergence of Network 2030 applications. There were many gaps highlighting use of AI models to overcome the identified deficiencies.

e) Changes in communication patterns can be observed as we saw more cases requiring group-based data sharing with high-precision guarantees. Even between two endpoints, the

need for multiple channels is observed. We can infer that finding mechanisms for efficient group communications will be of high significance.

f) Security and privacy are always high requirements. Due to lack of experts on this topic, it is not covered thoroughly. Instead a higher expectation from services such as high-precision, coordinated, qualitative, etc. have created a more perplexing situation in security for researchers on how to protect context of such services from in-network attacks.

g) Whether to consider fulfilling the gap independently or holistically can be inferred from the following data – by measuring weighted strength of each gap these gaps had values (strength) higher or equal to average (8), G.NS.1., G.NS.5., G.NS.11., G.NS.14., G.CAP.3., G.CAP.6., G.SEC.1., G.AIN2., G.NS.12., G.MGMT.5., G.SEC.3., G.NS.6., G.NS.15., G.CAP.4., G.NS.8., G.NS.9., G.NS.10. These gaps have research requirements in all factors and may be considered for holistic study.

9 Summary and conclusion

The focus group Network 2030 vision [1] aimed to provide a detail study of use cases and new services. A further assessment of those use cases, with respect to new 2030 services is presented in this technical report. Gaps were presented in three clauses pertaining to services (core network 2030), capabilities (continuous and new improvements), and new infrastructure (new verticals, AI, compute, etc.). The gap analysis gives a detailed description of state of the art vis-à-vis requirements of Network 2030 only from the functional aspect and no solution is proposed. Instead a further segmentation into factors (architectural, technology, protocol, and algorithms) where contributions would be necessary was provided.

Next steps will be to take a solution-based approach and validate them against these gaps. This report may also serve as a basis to evaluate any proposal on one topic (NCC, resilience, in-time services, etc.) and analyse its impact on other topics. Intuitively, all the functionality identified above together form Network 2030 and will co-exist. Therefore, a key aspect moving forward should be to devise and study any proposed solution's complexity, coverage, and impact on network behaviour through formal mechanisms.

Authors and contributors

This technical report collects the work and analysis from the original experts from the Network 2030 services and use cases reports.

Antonio Iera, DIMES, UNICAL

Claudia Campolo, UNIRC, Italy

Alex Clemm, Futurewei

Lijun Dong, Futurewei

Zhang Jingcheng

Fu Yuexia, China Mobile

Giacomo Morabito, UNICT

Kiran Makhijani, Futurewei, Ed.

Luigi Atzori, IEEE

Ning Wang, University of Surrey

Ren Yongmao

Xie Yunpeng, China Telecom

Yao Huijuan, China Mobile

Zhang Yuan, China Telecom

Appendix I

Table I.1 lists identified gaps.

Table I.1 – Table of identified gaps

Gap	Caption	Category	Arch	Proto	Algo	Tech
G.NS.1.	Quantifiable end-to-end latency	HPC		M	H	H
G.NS.2.	Service Constraints with varying bit rates	HPC		L	L	H
G.NS.3.	Application defined service customizations	HPC	L	M		
G.NS.4.	Pacing and queuing based on desired latency	HPC		L	H	M
G.NS.5.	In-network support for coordination	HPC	L	H	L	H
G.NS.6.	Application defined dependency	HPC	H	H	L	H
G.NS.7.	Support for dynamic pacing and feedback	HPC	M	L	L	H
G.NS.8.	Semantic- and context-based payload realization	QUAL	H	M	H	H
G.NS.9.	Application level packetization and encoding for qualitative payload	QUAL	H	M	H	H
G.NS.10.	Forwarding-node qualitative function	QUAL	H	M	H	H
G.NS.11.	Paced Synchronization and Prioritization	COORD	L	M	H	M
G.NS.12.	Application defined consistent throughput	COORD	L	M	H	H
G.NS.13.	Metadata for in-network transport feedback	COORD	M	M	L	L
G.NS.14.	On demand high-volume resource customization	VLV	L	M	M	H
G.NS.15.	Synchronization from multiple sources	VLV	M	H	H	M
G.NS.16.	Reliable high-speed data streaming	VLV	L	L	L	M
G.NS.17.	Cut-through application data forwarding with dynamic link resource	ABF	L	M	L	M
G.NS.18.	Congestion free cut-through burst transmission scheduling	ABF	M	L	M	M
G.CAP.1.	SLO-aware Network Service Interfaces	NSI	L	M	L	L
G.CAP.2.	SLO negotiation via Network Service Interfaces	NSI	L	M	L	L
G.CAP.3.	Support for coordinated services	NSI	L	M	H	M
G.CAP.4.	Support for qualitative services	NSI	M	M	H	H
G.CAP.5.	Inherent accounting support for SLOs	NSI	L	M	M	M
G.CAP.6.	Accommodation of network-level trust mechanisms	NSI	L	M	M	H
G.CAP.7.	Rapid Customization of Network Services	PROG	L	M	L	L
G.MGMT.1.	High-Precision Measurements	MGMT	L	M	M	M
G.MGMT.2.	Comprehensive device telemetry at scale	MGMT	L	L	L	M
G.MGMT.3.	Comprehensive end-to-end packet telemetry at scale	MGMT	L	M	L	L
G.MGMT.4.	Service level validation	MGMT	L	L	L	L

Table I.1 – Table of identified gaps

Gap	Caption	Category	Arch	Proto	Algo	Tech
G.MGMT.5.	Assessment of data or privacy leakage	MGMT	L	M	H	H
G.MGMT.6.	Human/Machine Interfaces for Intent	MGMT	L	M	L	L
G.MGMT.7.	Automated planning of explainable courses of actions for outcomes	MGMT	M	L	M	L
G.MGMT.8.	Decentralized and distributed architectures for low-latency management at scale	MGMT	M	M	M	L
G.SEC.1.	authorization of packets	SEC	L	L	H	H
G.SEC.2.	Authentication of packet headers and prevention of spoofing	SEC	L	M	M	M
G.SEC.3.	Homomorphic encryption	SEC	L	M	H	H
G.SEC.4.	Network 2030 Services Security: Coordinated Services	SEC	L	M	M	M
G.SEC.5.	Network 2030 Services Security: Qualitative Communications	SEC	L	M	M	M
G.RES.1.	SLO-aware resource protection	RES	L	M	M	M
G.RES.2.	SLO-aware path diversity and protection	RES	L	M	M	M
G.RES.3.	Resilience intent and levels in SLAs	RES	M	L	L	M
G.RES.4.	Resilience at system level	RES	M	L	M	M
G.RES.5.	Resilience control knobs	RES	L	L	M	M
G.LLS.1.	Cross-layer support	LLS	L	M	L	L
G.LLS.2.	ML techniques predicting congestion	LLS	L	L	L	M
G.LLS.3.	Application defined lossless-ness criteria	LLS	L	L	L	L
G.PRIV.1	Opacity of User data	PRIV	L	M	L	L
G.PRIV.2	Protecting service level parameters	PRIV	L	H	L	L
G.PRIV.3	Secured Storage	PRIV	L	L	M	L
G.PRIV.4	Flow anonymization	PRIV	L	H	L	L
G.VAL.1	Accuracy	VALID	L	L	M	M
G.VAL.2	Proof of Service Delivery	VALID	L	M	L	L
G.VAL.3	Accounting Advances	VALID	L	L	L	L
G.VAL.4	Accounting for new network services	VALID	L	L	L	L
G.NCC.1.	Computing-aware routing	NCC	M	L	M	L
G.NCC.2.	Coordination among different computing resources	NCC	L	M	M	L
G.NCC.3.	The reliability of computing resources	NCC		L	L	M
G.NCC.4.	Service requirements awareness	NCC		L	M	L
G.NCC.5.	Distributed and unified framework	NCC	M	M	L	M
G.ION.1.	Massive data computing, storage, collection and analytics	ION	M	L	M	L
G.ION.2.	High programmability	ION	L	L	L	L
G.ION.3.	Autonomy	ION	L	L	M	M

Table I.1 – Table of identified gaps

Gap	Caption	Category	Arch	Proto	Algo	Tech
G.MNS1.	Unified addressing/routing architecture	MNET	M	L	L	L
G.MNS2.	Integration of backbone and access functionalities	MNET	L	H	L	L
G.MNS3.	Quality of Service Support	MNET	L	M	M	M
G.MNS4.	Resource management	MNET	L	L	M	L
G.AIN1.	Addressing	AIN	L	M	L	L
G.AIN2.	Joint network, intelligence and computing orchestration	AIN	L	H	H	L
G.AIN3.	Group-based communications	AIN	M	H	L	L
G.SIOT.1.	Things virtualization	SIOT	M	L	M	L
G.SIOT.2.	Support of application-specific communication primitives	SIOT	L	L	M	L
G.SIOT.3.	Mobility of things and virtual entities	SIOT	M	M	L	L
G.SIOT.4.	Differentiated security and trust for resource constrained devices	SIOT	L	L	M	L

Appendix II

Table II.1 shows use case to service mapping.

Table II.1 – Table for use case to service mapping

SVC/UC	HPC	VLV	AI	NSI	TAC	MNET	LLS	REL	BDF	Compute
DT	H	L	H	H	H	L	H	H		L
TIRO	H	L		H	H	L	H			
NCC	L		H	H		L	L	M		H
SIoT	L		H			L		H		
ION	L	L	H			L		L	L	M
HSD	M	H		L		L	M	H	L	
ABF	M	M	H	L		L	M	L	M	L
HTC	M	H	L	H	L	L	L			L
STIN				M		H	L	L		L
IIoT		L			L	L	M	L		M