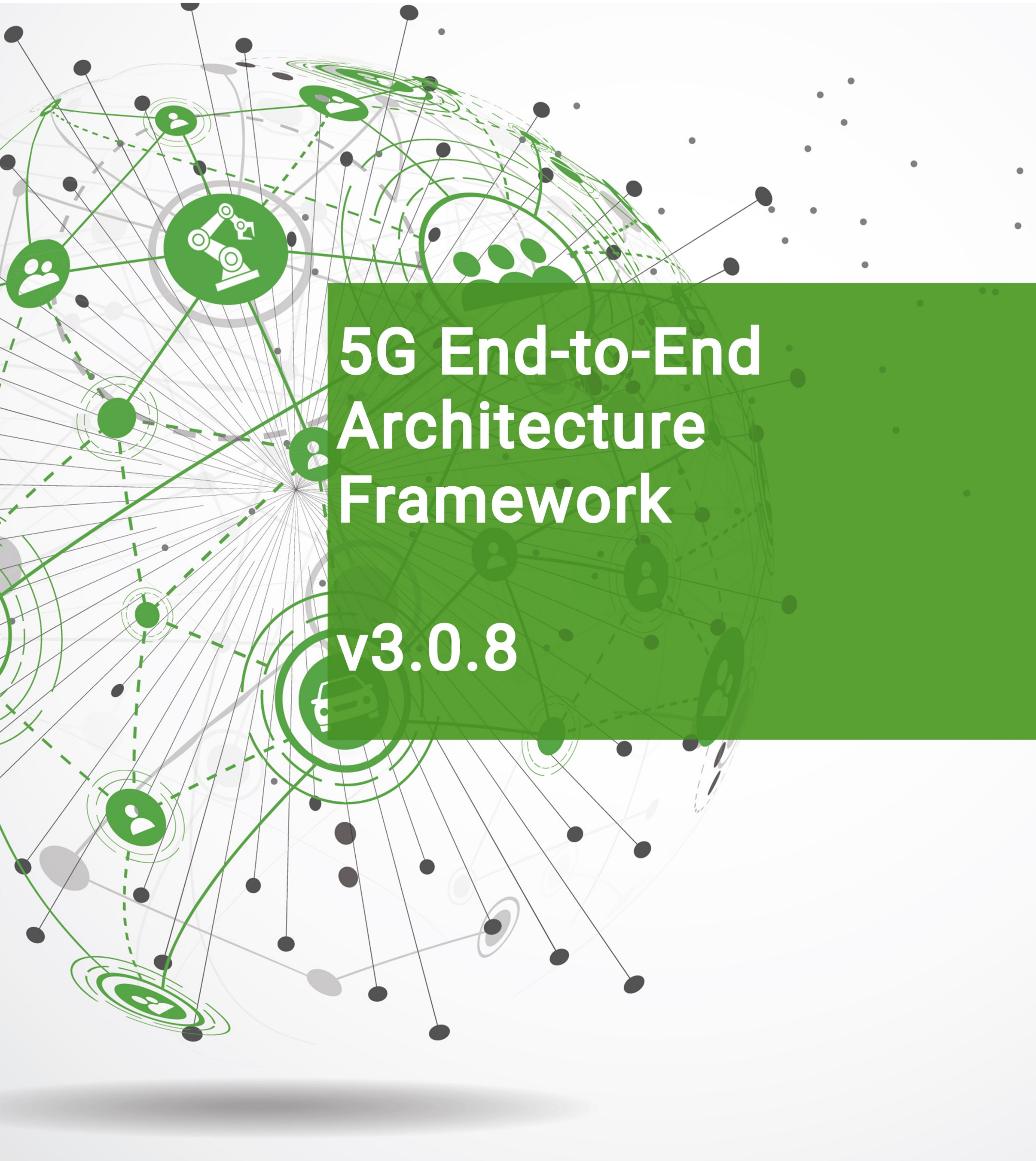
ngmn
the engine of
wireless innovation



5G End-to-End Architecture Framework

v3.0.8

5G End-to-End Architecture Framework

by NGMN Alliance

Version:	**3.0.8**
Date:	**28th-August-2019**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**

Project:	**End-to-End Architecture Framework**
Editor / Submitter:	**Sebastian Thalanany, U.S. Cellular**
Contributors:	**Srisakul Thakolsri (NTT DOCOMO), Adrian Neal (Vodafone), Steve Tsangkwong U (Orange), Richard Mackenzie (British Telecom), Peter Hedman (Ericsson), Paul Muschamp (British Telecom), Ahmed Alsohaily (Univ. Toronto), Chen Wei (China Mobile) Dan Wang (China Mobile), Shahar Steiff (PCCW Global), Hans J. Einsiedler (Deutsche Telekom), Tayeb Benmeriem (Orange), Ines Riedel (Amdocs), Philipp Deibert (NGMN), Javan Erfanian (Bell Canada), Farooq Bari (AT&T), Jan Groenendijk (Ericsson), , Min Zuo (China Mobile), Charles Hartmann (Orange), Paul Bradley(Gemalto), Steve Babbage (Vodafone), Marc Kneppers (Telus), Axel Nennker (Deutsche Telekom), Jovan Golic (Telecom Italia), Dirk v. Hugo (Deutsche Telekom), Gerhard Kadel Deutsche Telekom), Gerd Zimmermann (Deutsche Telekom), Serge Manning (Sprint), Gary Li (Intel)**
Approved by / Date:	**NGMN Board, 13th September 2019**

The information contained in this document represents the current view held by NGMN e.V. on the issues discussed as of the date of publication. This document is provided "as is" with no warranties whatsoever including any warranty of merchantability, non-infringement, or fitness for any particular purpose. All liability (including liability for infringement of any property rights) relating to the use of information in this document is disclaimed. No license, express or implied, to any intellectual property rights are granted herein. This document is distributed for informational purposes only and is subject to change without notice. Readers should not design products based on this document.

Abstract: Short introduction and purpose of document

This document delineates the requirements in terms of entities and functions that characterise the capabilities of an E2E (end-to-end) framework. Architectural perspectives and considerations associated with the service categories - eMBB, mIoT, URLLC - envisioned for 5G (Fifth Generation) underscore the delineation of the E2E framework requirements. These requirements are intended as guidance in the development of inter-operable and market enabling specifications for a 5G ecosystem,

Document History

Date	Version	Author	Changes
14/09/2016	V 0.1.0	Adrian Neal, Vodafone	First version
20/09.2016	V 0.2.0	Adrian Neal, Vodafone	Addition of Devices section
04/10/2016	V 0.3.0	Sebastian Thalanany, U.S. Cellular, Adrian Neal, Vodafone	Addition of text to Introduction and References sections
11/10/2016	V 0.3.1	Sebastian Thalanany, U.S Cellular, Adrian Neal, Vodafone, Richard Mackenzie, British Telecom.	Text in the Devices section.
06/12/2016	V 0.3.2	Adrian Neal, Vodafone, Sebastian Thalanany, U.S. Cellular.	Addition of text agreed as a liaison to 3GPP SA2, and in definitions and devices sections.
16/01/2017	V 0.4.0	Adrian Neal, Vodafone	Addition of agreed text to Sections 5 (Network Slicing) and 9 (Management and Orchestration).
27/01/2017	V 0.5.0	Adrian Neal, Vodafone	Addition of agreed text to Sections 4.2, 6.1, 6.1.5, 6.1.6, 6.2.2, 7.1, 8.1, 9.1, 11.1, 11.2, 13.1, and 15.
28/01/2017	V 0.5.1	Sebastian Thalanany, U.S. Cellular	Addition/revision of text in sections covering definitions, access, and core.
31/01/2017	V 0.5.2	Sebastian Thalanany, U.S. Cellular	Added document purpose, Sections 6.1 and 6.2 to support ultra-low latency, high reliability and availability usage scenarios.
06/02/2017	V 0.5.3	Sebastian Thalanany, U.S. Cellular, Ahmed Alsohaily, University of Toronto	Definitions for network slice blueprint, network slice instance, service instance, and network function have been provided. Revised NSP, SP, and added VNSP. Revised the Abstract and clarified abbreviations. Filled up Section 6.3, added reference [8] and provided minor edits in sections 3-6, 8, 10 and 12.
28/02/2017	V 0.5.4	Sebastian Thalanany, U.S. Cellular, Chen Wei, CMCC, Paul	Updated text, based on comments from the call on February 23rd,

		Muschamp, BT, Ahmed Alsohaily, University of Toronto	2017. Added Service-Based Architecture (Section 6.1), 5G RAN functional decomposition (Sections 6.3.1 and 11.1), Reference [12]].
13/03/2017	V 0.5.5	Sebastian Thalanany, U.S. Cellular	Qualified “minimised coupling” in Section 6.1. Added comment in Section 6.3.1, indicating an elevation of the requirements text, while not alluding to SDO options, or implementation.
28/03/2017	V 0.6.0	Adrian Neal, Vodafone.	Inclusion of agreed text from 23rd March call and contributions. Outstanding comments converted to Editor’s Notes.
06/04/2017	V 0.6.1	Adrian Neal, Vodafone.	Inclusion of agreed text to Sections 6.3.1.1, 6.4.4, 7.1-4, 8.1-5 and 11.3.
28/04/2017	V 0.6.2	Hans J. Einsiedler, DTAG, Shahar Steiff, PCCW Global, Sebastian Thalanany, U.S. Cellular, Adrian Neal, Vodafone	Editorial scrub Minor additions in 6.1, addition of 9.4 Federated Orchestration, added content in the General sections Section 6.2.1 Introduced the concept of “Microservices” as an enabling facet in the end-to-end framework, in Section 6.2.3. Addition of agreed content from 27th April conference call.
01/05/2017	V 0.6.3	Adrian Neal, Vodafone	Additional changes to Section 9.4
04/05/2017	V 0.6.4	Adrian Neal, Vodafone, Tayeb Benmeriem, Orange, Ines Riedel, Amdocs, Philipp Deibert, NGMN Office	Agreed changes to Sections 2, 6.1, 9.2, 9.3, 9.4, 9.5 and 15 from the 4th May conference call.
11/05/2017	V 0.6.5	Klaus Moschner, NGMN, Adrian Neal, Vodafone	Changes to Section on Federated Orchestration and renaming MANO to 5G E2E MANO. Addition of Bell Canada as contributor (from V0.6.4).
04/06/2017	V 0.6.6	Sebastian Thalanany, U.S. Cellular	Edited Section 9, and 9.1. Added E2E (End-to-End) definition Added text and diagrams – network slice management and orchestration, in Sections 9.2, 9.2.1, 9.2.2, and 9.2.3
07/06/2017	V 0.6.7	Sebastian Thalanany, U.S. Cellular	Updated E2E related definition. Included clarification text in Sections 9.2.1, 9.2.2, and 9.2.3

19/06/2017	V 0.6.8	Sebastian Thalanany, U.S. Cellular	Added and revised text in Section 5.1.
27/06/2017	V 0.6.9	Adrian Neal, Vodafone	Agreed updates to Sections 5.1, 9.1 from 22nd June 2017 conference call.
11/07/2017	V 0.7.0	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange	Updated diagrams and text in Sections 4.2, 6.2.2, 9.2.4, 9.2.5, 9.2.6, 9.2.7, 9.2.8. and 9.2.9
13/07/2017	V 0.7.1	Sebastian Thalanany, U.S. Cellular	Merged updates from V 0.6.9r1.
21/07/2017	V 0.7.2	Adrian Neal, Vodafone	Inclusion of changes to Sections 1, 6.1.1, 6.1.2 and 6.2.2, and new Figure 2, from July 20th conference call.
27/07/2017	V 0.7.3	Adrian Neal, Vodafone	Editorial clean up agreed on 27th July conference call.
22/08/2017	V 0.7.6	Sebastian Thalanany, U.S. Cellular, NGMN Security Competence Team (SCT): especially, Min Zuo, China Mobile, Charles Hartmann, Orange, Axel Nennker, Deutsche Telekom, Paul Bradley, Gemalto, and Jovan Golic, Telecom Italia.	Merged agreements through interim revisions post August 3, August 10, 2017 reviews and calls. Incorporated NGMN Network Management and Orchestration team inputs on management and orchestration. Incorporated NGMN Security Competence Team input on the security and identity management sections resulting from several iterations, discussions and calls. Confining devices to customer-side, “unified customer database” removed. Editorial changes together with diagrams, references, and cross-reference updates.
01/09/2017	V0.7.9	Adrian Neal, Vodafone, Steve Babbage, Vodafone, SCT NGMN: especially, Charles Hartmann, Orange, Paul Bradley, Gemalto, and Jovan Golic, Telecom Italia, Alec Brusilovsky, Interdigital.	Agreements from 31st August call. Restructuring of Security and Identity management sections. Re-definition of devices as Endpoint/User devices.
03/09/2017	V0.7.10	Adrian Neal, Vodafone, Hans Einsiedler, Deutsche Telekom, Sebastian Thalanany, U.S. Cellular, Peter Hedman, Ericsson, Srisakul Thakolsri, NTT DOCOMO.	Update to Figure 1, Clarification of “VF” in Figure 2, further clarifications to Sections 5, 6, 8, 9, and 14 and new Abbreviations plus a new definition for X-Haul. All arising from the internal company review period and contributions to the 1st September conference call.
07/09/2017	V0.7.11	Adrian Neal, Vodafone, Peter Hedman, Ericsson, Srisakul	Agreements from the 7th September E2E Architecture

		Thakolsri, NTT DOCOMO, Farooq Bari, AT&T, Shahar Steiff, PCCW Global, Paul Muschamp, BT.	Framework conference call. Revisions to Sections 6.4.3, 7.3, 11.2.3 and 16.
11/09/2017	V0.7.12	Shahar Steiff, PCCW Global, NGMN SCT: Jovan Golic, Telecom Italia, Steve Babbage, Vodafone, Charles Hartmann, Orange, Paul Bradley, Gemalto, Alec Brusilovsky, InterDigital.	Rewording in Section 11.2.3. Detailed revision of Sections 13 and 15 agreed on the 7th September SCT conference call.
15/09/2017	V0.8.0	Srisakul Thakolsri, NTT DOCOMO, Axel Nennker, Deutsche Telekom, Charles Hartmann, Orange, Adrian Neal, Vodafone.	Agreed changes to Section 9.2 from the 14th September conference call. Editorial clean-up including realignment of reference numbering and editorial comments by e-mail.
17/09/2017	V0.8.1	Adrian Neal, Vodafone.	Removal of two references not used in the text. Preparation for Board Approval.
09/02/2018	V1.1.0	SCT, NGMN: especially, Jovan Golic, Telecom Italia, Steve Babbage, Vodafone, Charles Hartmann, Orange, Paul Bradley, Gemalto, Marc Kneppers, Telus, Yogendra Shah, InterDigital	Sections 13.6 on security challenges and 15.1 on IDM open issues from V0.8.1 are resolved, updated and incorporated in Sections 13.1-13-5 and 15.1, respectively. Seven new references [24]-[31] are introduced.
12/02/2018	V2.0.0	Adrian Neal, Vodafone	Editorial clean up and preparation for Board Approval after agreement from the joint E2E Arch/SCT conference call on 12th February.
21/06/2018	V2.0.1	Dirk von-Hugo, Deutsche Telekom	Added text on Fixed-Mobile convergence
11/07/2018	V2.0.2	Hans J. Einsiedler, Gerhard Kadel, Dirk v. Hugo, DTAG, Deutsche Telekom	Update on FMC
02/08/2018	V2.0.2	Sebastian Thalanany, U.S. Cellular	Editorial updates in the FMC section.
03/08/2018	V2.0.3	Hans J. Einsiedler, Gerhard Kadel, Dirk von Hugo, Gerd Zimmermann, Deutsche Telekom	Added and detailed description of Chapter 6.1.2 Updated Figure 3, current 4, 5, 6
17/08/2018	V2.0.3	Hans J. Einsiedler	Updated Figure 3 and additional explanations
28/08/2018	V2.0.3	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange	Added content in sections 6.3, 6.3.1, and 6.3.2. Added placeholder in section 6.4, and 9.3 Editorial scrub.

12/09/2018	V2.0.4	Farooq Bari, AT&T, Sebastian Thalanany, U.S. Cellular	Text added in section 6.2.1 to avoid tunnels for certain use cases associated with 5G IoT. Editorial scrub.
13/09/2018	V2.0.4	Tayeb Benmeriem, Orange, Sebastian Thalanany, U.S. Cellular	Input from on autonomic networking, section 6.4. Editorial scrub.
27/09/2018	V2.0.4	Serge Manning, Sprint, Adrian Neal, Vodafone	Merged input on E2E network automation requirements, based on zero-touch management automation, presented on the September 20th E2E call.
07/10/2018	V2.0.4	Sebastian Thalanany, U.S. Cellular	Merged version containing all agreed draft inputs from the call on September 20, 2018. Added missing cross-reference links, and updated stale cross-reference links. Added requirements qualifiers in section 6.1 “information hiding”, and “high cohesion” corresponding to encapsulation to minimize system impact because of software building block changes, and to group similar building blocks together for robust and reliable behaviors. Editorial scrub
25/10/2018	V2.0.4	Tayeb Benmeriem, Orange, Adrian Neal, Vodafone	New section 6.4.2 created on federated E2E decision-making across autonomic domains. Updated with clarifications on network slicing aspects.
29/10/2018	V2.0.5	Sebastian Thalanany, U.S. Cellular	Draft working baseline for further review and update, after the conference call on October 25, 2018. Editorial scrub.
20/11/2018	V2.0.5	Adrian Neal, Vodafone	Updates to multi-access edge computing text and reference in section 6.1. Editorial scrub.

11/12/2018	V2.0.5	Sebastian Thalanany, U.S. Cellular	Replied to comments and consolidated updates from the previous E2E call.
17/12/2018	V2.0.5	Tayeb Benmeriem, Orange	Inputs for section 6.3, and section 6.4, with respect to DLT and Autonomic networking, consolidated with previous inputs into version 1.1.5
29/01/2019	V2.0.6	Sebastian Thalanany, U.S. Cellular	Updated document in terms of reviews from the last call, together with a derivation of requirements, based on previous inputs from Tayeb, in sections 6.3 and 6.4. Editorial scrub.
10/02/2019	V2.0.6	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange	Updated to include review comments, from the call on 31/1/2019. Added the motivators for autonomic networking to meet the diverse demands of 5G usage scenarios. Updated pointers to references.
06/03/2019	V2.0.7	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange	Updated to include review comments, from the call on 21/2/2019, including inputs from Tayeb. Elaborated and created a separate section on usage scenarios, under section 6.4 on AuN (Autonomic Networking). Re-arranged section 6.4 on AuN for condensing and clarifying the various system-wide implications to serve the unique and diverse E2E demands of a 5G ecosystem. Updated section 8.3, with input from Fran O'Brien. Updated pointers to references.
22/03/2019	V2.0.8	Sebastian Thalanany, U.S. Cellular	Updated and condensed section 6.4 on AuN (Autonomic Networking). Added section 17 (Annex for AuN containing additional information).
28/03/2019	V2.0.9	Tayeb Benmeriem, Orange, Adrian Neal, Vodafone, Srisakul Thakolsri, NTT DoCoMo, Farooq	Consolidated review comments, from the March 28th, 2019 call.

		Bari, AT&T, Sebastian Thalanany, U.S. Cellular.	
3/04/2019	V2.1.0	Sebastian Thalanany, U.S. Cellular.	Updated the document based on the consolidated review comments, from the March 28th, 2019 call. Revised the figures in section 6.4 on autonomic networking, cross-references, the definition section, and resolved review comments.
12/04/2019	V2.1.1	Srisakul Thakolsri, NTT DoCoMo	Editorial clean-up and putting back some texts related to the section 13.5 that were missing from the version 2.0
16/04/2019	V3.0.0	Sebastian Thalanany, U.S. Cellular, Srisakul Thakolsri, NTT DoCoMo, Tayeb Benmeriem, Orange	Updated the document with a resolution of all existing review comments. Added a reference to the vision of the NGMN white paper on 5G, pertaining to the requirements in section 6.4 Editorial scrub to resolve formatting errors cross-references.
26/05/2019	V3.0.1	Dan Wang, China Mobile	Introduction of User Plane Function services in section 6.6.4
21/06/2019	V3.0.2	Sebastian Thalanany, U.S. Cellular, Gary Li, Intel	Accepted all comments, after review, in section 6.6.4. Editorial scrub.
25/06/2019	V3.0.3	Adrian Neal, Vodafone, Sebastian Thalanany, U.S. Cellular	No content changes. Edited the colour of a segment of text in section 13.1, from red to black, which was left over from a previous review and acceptance.
09/07/2019	V3.0.4	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange	Added two new subsections: 1 AI (Artificial Intelligence) and ML (Machine Learning) in AMC 2 KP (Knowledge Plane) use cases for AMC for mobile and fixed access scenarios.
24/07/2019	V3.0.5	Tayeb Benmeriem, Orange, Sebastian Thalanany, U.S. Cellular	Added content in the following sections: 6.4.5 AI and ML in AMC 6.4.6 AMC KP use cases
06/08/2019	V3.0.6	Tayeb Benmeriem, Orange, Sebastian Thalanany, U.S. Cellular	Content revision, derivation of principles, comment resolution from the previous call in the following sections: 6.4.5 AI and ML in AMC 6.4.6 AMC KP use cases

			Editorial scrub for document baseline establishment.
19/08/2019	V3.0.7	Sebastian Thalanany, U.S. Cellular, Tayeb Benmeriem, Orange, Srisakul Thakolsri, NTT DoCoMo, Gary Li, Intel	Baseline version – E2E Phase-2 Comment resolution. Editorial scrub.
27/08/2019	V3.0.8	Sebastian Thalanany, U.S. Cellular, Srisakul Thakolsri, NTT DOCOMO, Gary Li, Intel	Editorial clean up

1 INTRODUCTION

The purpose of this document is to provide a high-level framework of architecture principles and requirements that provide guidance and direction for NGMN partners and standards development organisations in the shaping of the 5G suite of interoperable capabilities, enablers, and services. It builds on the architectural concepts and proposals implied by the NGMN White Paper [1] and subsequent deliverables published by NGMN. It is anticipated that this document will have versions, beyond an initial version, to reflect additional forward-looking requirements and/or updates as needed.

The elements of functional virtualisation shift of computing to the edges of the network, and leveraging of spectrum distribution and flexibility, are among the dominant themes that shape the 5G ecosystem [1]. Optimisation of operational and performance efficiencies, while creating and delivering an exceptional and customisable user experience is of paramount significance [2][3].

2 REFERENCES

[1] NGMN 5G White Paper v1.0, Feb. 2015.
[2] 3GPP TR 22.891: Study on New Services and Markets Technology Enablers, Release 14, v1.0.0, Sep. 2015.
[3] Recommendations for NGMN KPIs and Requirements for 5G, June 2016.
[4] NGMN Description of Network Slicing Concept v1.0.8, Sep. 2016.
[5] E.U. 5G-PPP project TRANSFORMER. http://5g-transformer.eu/
[6] NGMN 5G security recommendations Package #2: Network Slicing, Apr. 2016. https://www.ngmn.org/uploads/media/160429_NGMN_5G_Security_Network_Slicing_v1_0.pdf
[7] ITU-T, “The tactile internet,” ITU-T technology watch report, Aug. 2014.
[8] NGMN 5G security recommendations Package #3: Mobile Edge Computing / Low Latency / Consistent User Experience, Oct. 2016. https://www.ngmn.org/uploads/media/161028_NGMN-5G_Security_MEC_ConsistentUExp_v1.3_final.pdf
[9] NGMN Perspectives on Vertical Industries and Implications for 5G v2.0, Sep. 2016.
[10] Dmitry Namiot, Manfred Sneps-Sneppe, “On Microservices Architecture”, International Journal of Open Information Technologies ISSN: 2307-8162 vol. 2, no. 9, 2014, pp. 24-27.
[11] NGMN Project RAN Evolution: Multi-RAT Joint Radio Operation (MRJRO) v1.1, Mar. 2015.
[12] NGMN Project RAN Evolution: Further Study on Critical C-RAN Technologies v1.0, Mar. 2015.
[13] NGMN “5G Network and Service Management, including Orchestration” v2.12.6. Mar. 2017.
[14] NGMN 5G security recommendations Package #1, May 2016.
[15] General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679), http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf
[16] Regulation on Privacy and Electronic Communications, https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications
[17] ETSI NFV-SEC 003: Security and Trust Guidance, Dec. 2014.
[18] ETSI NFV-SEC 009: Report on use cases and technical approaches for multi-layer host administration, Jan. 2017.
[19] ETSI NFV-SEC 012: System architecture specification for execution of sensitive NFV components, Jan. 2017.
[20] ETSI NFV-SEC-013: Security Management and Monitoring Specification, Feb. 2017.
[21] ETSI NFV-SEC 014: Security Specification for MANO Components and Reference points, May 2017.
[22] FIPS PUB 140-2: Security Requirements for Cryptographic Modules, available at http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf
[23] FIDO alliance, https://fidoalliance.org/
[24] Z. Kotulski, T. Nowak, M. Sepczuk, M. Tunia, R. Artych, K. Bocianiak, T. Osko, and J.-P. Wary, “On end-to-end approach for slice isolation in 5G networks – Fundamental challenges,”

Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 783-792, Prague, 3-6 Sep. 2017.

[25] NIST.IR 8114: Report on Lightweight Cryptography, Mar. 2017 *https://www.nist.gov/programs-projects/lightweight-cryptography*

[26] GSMA, "Identity and Access Management Requirements, version 1.0," Nov. 13, 2017.

[27] ISO/IEC 29192: Lightweight Cryptography, 2012-.

[28] ISO/IEC 29167: Automatic Identification and Data Capture Techniques, 2014-.

[29] Ashutosh Dutta, "Security Challenges and Opportunities in SDN/NFV and 5G Networks," AT&T presentation, June 15, 2017 https://docbox.etsi.org/Workshop/2017/201706_SECURITYWEEK/06_5GSECURITY/S01/AT%26T_Dutta.pdf

[30] ETSI NFV-SEC 002: "Cataloguing Security Features in Management Software," Aug. 2015.

[31] 3GPP TS 33.501: "Security Architecture and Procedures for 5G System," Release 15, v0.6.0, Dec. 2017.

[32] BBF SD-407: "5G Fixed Mobile Convergence Study," CONTRIB-20329 FMC Architectural Models, Nov. 2017

[33] BBF SD-420: "5G Fixed Mobile Convergence Study (External), work in progress," January 2018

[34] BBF TR-348: "Hybrid Access Broadband Network Architecture," July 2016 https://www.broadband-forum.org/technical/download/TR-348.pdf

[35] 3GPP TR 23.716 V0.5.0, "Study on the Wireless and Wireline Convergence for the 5G system architecture (Release 16)," work in progress, June 2018

[36] 3GPP TR 23.793 V0.5.0, "Study on Access Traffic Steering, Switching and Splitting support in the 5G system architecture", (Release 16), work in progress, June 2018

[37] ITU-T: "Distributed Ledger Technologies and Financial Inclusion," ITU-T Focus Group Digital Financial Services, Focus Group Technical Report, March 2017.

[38] IETF: "Co-existence of 3GPP 5GS and Identifier Locator Separation Architecture," draft-homma-dmm-5gs-id-loc-coexistence-01, May 15, 2018

[39] 3GPP TR 23.724 V1.0.0, "Study on Cellular IoT support and evolution for the 5G System", (Release 16) work-in-progress, September 2018

[40] TM Forum TR279: "CSP Use Cases Utilizing Blockchain R17.5.0", March 2018

[41] ETSI White Paper No.16: GANA (Generic Autonomic Networking Architecture) (http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp16_gana_Ed1_20161011.pdf)

[42] ETSI TS 103 195-2: "Autonomic network engineering for the self-managing Future Internet (AFI); Generic Autonomic Network Architecture", (http://www.etsi.org/deliver/etsi_ts/103100_103199/10319502/01.01.01_60/ts_10319502v010101p.pdf)

[43] ETSI : "Zero Touch Network and Service Management (ZSM); Means of Automation," DRAFT ETSI GR ZSM-005 V0.1.0 (2018-10)

[44] ONAP Architecture Overview: (https://www.onap.org/wp-content/uploads/sites/20/2018/06/ONAP_CaseSolution_Architecture_0618FNL.pdf)

[45] ETSI White Paper No. 28: " MEC (Multi-access Edge Computing) in 5G", First edition, June 2018,(https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf)

[46] RFC7575: "Autonomic Networking: Definitions and Design Goals", IRTF, June 2015.

[47] ETSI 5G PoC White Paper No.1: C-SON Evolution for 5G, Hybrid SON Mappings to the ETSI GANA Model, and achieving E2E Autonomic (Closed-Loop) Service Assurance for 5G Network Slices by Cross-Domain Federated GANA Knowledge Planes (https://intwiki.etsi.org/index.php?title=Accepted_PoC_proposals)

[48] ETSI TR 103 473 V1.1.2 on Autonomicity and Self-Management in the Broadband Forum (BBF) Architectures (a GANA instantiation onto BBF architecture scenarios)

[49] ETSI TR 103 404 on Autonomicity and Self-Management in the Backhaul and Core network parts of the 3GPP Architecture (a GANA instantiation onto Backhaul and EPC Core Network)

[50] NIST IR (Internal Report) 82 92 title is “Blockchain Technology Overview”, October 2018, Dylan Yaga, Peter Mell, Nik Roby, Karen Scarfone

[51] TM Forum IG 1175, R18.5, “Trust & Traceability in Digital Ecosystems”, https://www.tmforum.org/resources/exploratory-report/ig1175-trust-and-traceability-in-digital-ecosystems-r18-5-0/

[52] 5G-PPP Cognet, http://www.cognet.5g-ppp.eu/wp-content/uploads/2015/08/CogNet_D21_v09_Final.pdf

[53] ITU-T Recommendation Y 3324 “Requirements and Architectural Framework for Autonomic Management and Control of IMT-2020 Networks

[54] Mitchell, T., “ Machine Learning”, McGraw Hill, 1997

[55] J. Pérez-Romero, O. Sallent, R. Ferrús, R. Agustí, “Knowledge-based 5G Radio Access Network Planning and Optimization”, The Thirteenth International Symposium on Wireless Communication Systems (ISWCS-2016), Poznan, Poland, September 2016.

[56] ETSI GANA Model in 5G Network Slicing: PoC by ETSI TC INT/ AFI (Autonomic Future Internet), “WG C-SON Evolution for 5G, Hybrid SON Mappings to the ETSI GANA Model, and achieving E2E Autonomic (Closed-Loop) Service Assurance for 5G Network Slices by Cross-Domain Federated GANA Knowledge Planes”, October 25th, 2018

[57] [57] White Paper No.4 of the ETSI TC INT / 5G PoC on “ETSI GANA as Multi-Layer Artificial Intelligence (AI) Framework for Implementing AI Models for Autonomic Management & Control (AMC) of Networks and Services; and Intent-Based Networking (IBN) via GANA Knowledge Planes”

3 DEFINITIONS

AuF
Autonomic Function, which is a type of function that does not require configuration, (except for being subject to input governance policies and the setting of operational mode (Open loop or Closed loop) and is able to derive all the necessary information, through self-knowledge, discovery, or policies.

E2E
End-to-End, which refers to communications between two endpoint devices or user equipment, across any arrangement of intervening administrative domains

DLT
Distributed Ledger Technology. This is a distributed database that leverages the blockchain framework, for storage, access, and adding data records securely, by authorized entities.

Haptic Sense
Haptic sense is perception characterised by touch. This type of perception is associated with tactile sense (derived from the Latin: *Tangere* - to touch), and kinaesthetic sense (derived from the Greek: *Kinesis* – movement, and *Aesthesis* – perception), for example body movement.

Network Function (NF)
Processing functions in a network. This includes a variety of control plane, user plane, and service functions that span the layers of the protocol stack. (e.g. radio network functions, physical layer functions, Internet Protocol (IP) routing functions, applications etc.) [4].

Network Service Provider (NSP)
Entity that provides network access service and owns related resources and functions (e.g. virtualised or physical) for providing network access. The resources and functions include spectrum, mobility and access

management across heterogeneous and/or composite access networks, network management and orchestration, and network elements.

Network Slice Blueprint (NSB) A complete description of the structure, configuration and the plans/work flow for how to instantiate and control the Network Slice Instance during its life cycle. A Network Slice Blueprint enables the instantiation of a Network Slice, which provides certain network characteristics (e.g. ultra-low latency, ultra-reliability, value-added services for enterprises, etc.). A Network Slice Blueprint refers to required physical and logical resources and/or to Sub-network Blueprint(s) [4].

Network Slice Instance (NSI) A set of run-time network functions, along with physical and logical resources to run these network functions, forming a complete instantiated logical network to meet certain network characteristics required by the Service Instance(s). A network slice instance may be fully or partly, logically and/or physically, isolated from another network slice instance [4].

Proprioceptive Sense Proprioceptive sense is perception characterised by a combination of body position and movement. This type of perception pertains to stimuli that are sensed and generated within an organism.

Service Instance (SI) An instance is a run-time construct of an end-user service or a business service that is realised within or by a Network Slice [4].

Service Provider (SP) Entity that provides an application layer service. The entity may be a third-party, or an NSP.

Vestibular Sense Vestibular sense is perception characterised by balance. This type of perception pertains to sensing via a cavity or vestibule, typically associated with the inner ear, which affects the state of balance of the body.

Visual Sense Visual sense is ocular perception that characterises seeing. This type of perception pertains to sensing via the eye.

X-Haul A common flexible transport solution for future 5G access networks, which aims to integrate fronthaul and backhaul networks with all their wired and wireless technologies in a common packet-based transport network under SDN-based (software defined networks) and NFV-enabled (network functions virtualization) common control.

4 HIGH LEVEL END-TO-END ARCHITECTURE

4.1 Background

4.2 High level architecture

NGMN envisions an architecture that leverages the structural separation of hardware and software, as well as the programmability offered by Software Defined Networks (SDNs) and Network Function Virtualisation (NFV). As such, the 5G architecture is a native SDN/ NFV architecture covering aspects ranging from endpoint/user equipment, (mobile/ fixed) infrastructure, network functions, value enabling capabilities and all the management functions to orchestrate the 5G system. Application Program Interfaces (APIs) are provided on the relevant reference points to support multiple use cases, value creation and business models.

The architecture includes layers above the network layer. It allows for federation between separately administered domains at the resource and service layers to realise end-to-end network and service slice instances where one or more service providers or network service providers are involved. This naturally implies federation of network management and service orchestration as well.

The overall arrangement of actors, in a virtualised framework for service orchestration, utilising network slicing as a foundational building block, in the context of one or multiple administrative domains, broadly referred to as domains is depicted in Fig.1.

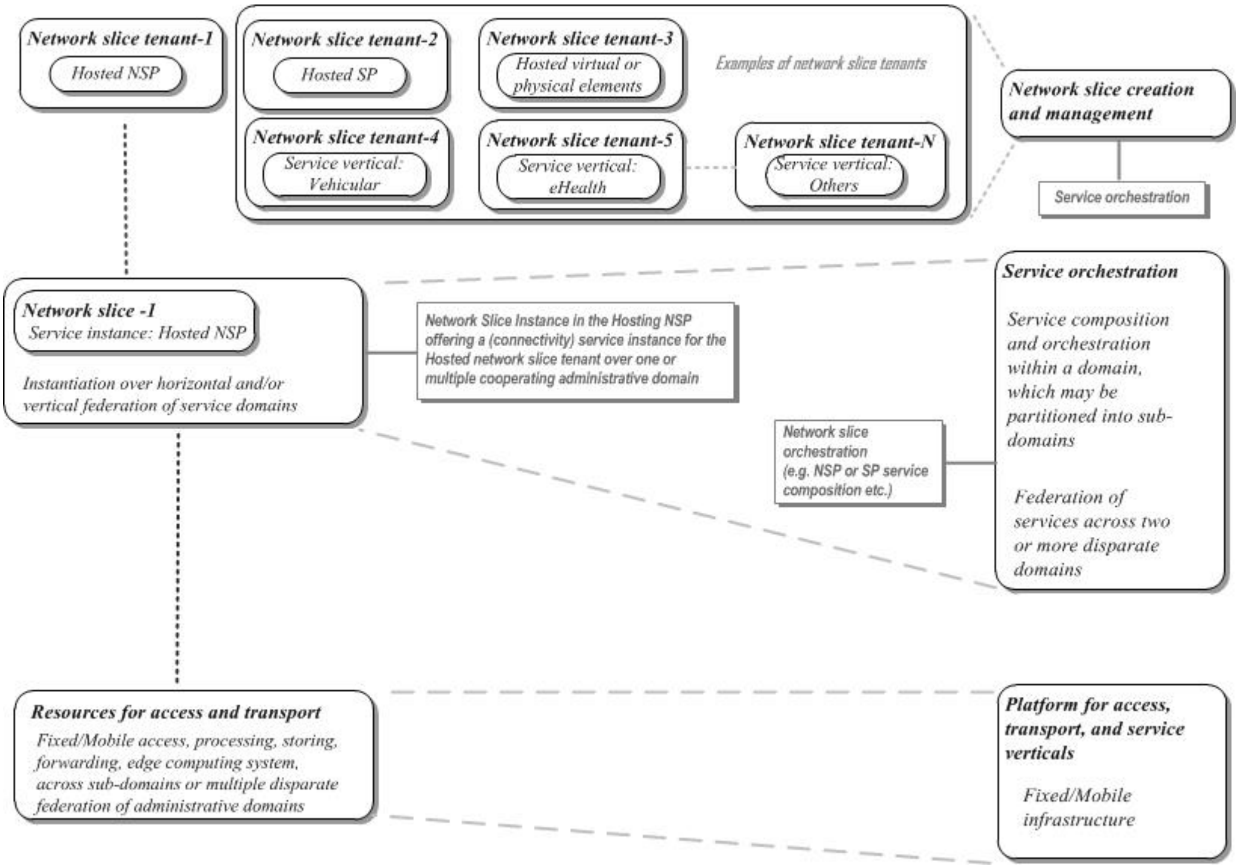

Network slice tenant-1
Hosted NSP
Network slice tenant-2
Hosted SP
Network slice tenant-3
Hosted virtual or physical elements
Examples of network slice tenants
Network slice tenant-4
Service vertical: Vehicular
Network slice tenant-5
Service vertical: eHealth
Network slice tenant-N
Service vertical: Others
Network slice creation and management
Service orchestration
Network slice -1
Service instance: Hosted NSP
Instantiation over horizontal and/or vertical federation of service domains
Network Slice Instance in the Hosting NSP offering a (connectivity) service instance for the Hosted network slice tenant over one or multiple cooperating administrative domain
Network slice orchestration (e.g. NSP or SP service composition etc.)
Service orchestration
Service composition and orchestration within a domain, which may be partitioned into sub-domains
Federation of services across two or more disparate domains
Resources for access and transport
Fixed/Mobile access, processing, storing, forwarding, edge computing system, across sub-domains or multiple disparate federation of administrative domains
Platform for access, transport, and service verticals
Fixed/Mobile infrastructure

Fig.1. NGMN End-to-End 5G Framework vision.

Fig.2. presents an alternative view, depicting how federation of resources and services occurs between different administrative domains to provide the end-to-end service. Further detail is elaborated in the following sections of this document.

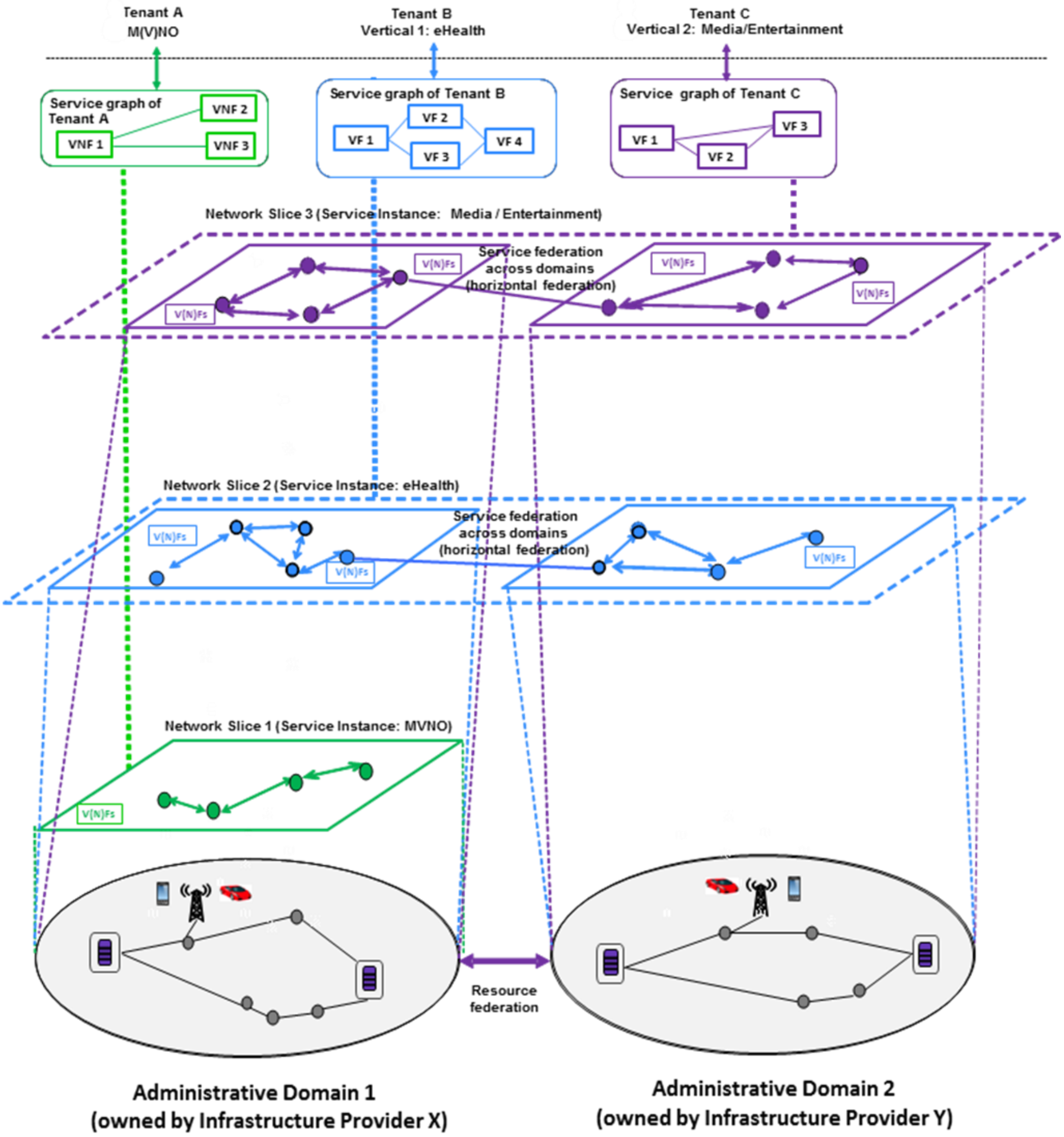

Tenant A
M(V)NO
Tenant B
Vertical 1: eHealth
Tenant C
Vertical 2: Media/Entertainment
Service graph of Tenant A
VNF 1
VNF 2
VNF 3
Service graph of Tenant B
VF 1
VF 2
VF 3
VF 4
Service graph of Tenant C
VF 1
VF 2
VF 3
Network Slice 3 (Service Instance: Media / Entertainment)
Service federation across domains (horizontal federation)
V(N)Fs
Network Slice 2 (Service Instance: eHealth)
Service federation across domains (horizontal federation)
Network Slice 1 (Service Instance: MVNO)
Resource federation
Administrative Domain 1 (owned by Infrastructure Provider X)
Administrative Domain 2 (owned by Infrastructure Provider Y)

Fig.2. Resource and service federation across administrative domains

The European Union’s Horizon 2020 5G-PPP phase II project 5G-TRANSFORMER [5] has adopted the NGMN vision. The term VF refers to virtualised service functions at layers above the network layer.

5 NETWORK SLICING

5.1 General

The scope of a network slice is end-to-end. The 5G network shall be capable of slicing by service categories that consist of enhanced Mobile Broadband (eMBB, massive Internet of Things (mIoT), Ultra-Reliable Low-Latency Communication (URLLC), and other new arising categories.

The behaviour of a network slice, in terms of relevant resources, virtualised entities and functions, non-virtualised entities, at the user plane, control plane, and management/orchestration plane, is realised via the composition and instantiation of a network slice.

From an administrative domain perspective an end-to-end network slice may be within a domain, sub-domain, or across domains.

An end-to-end network slice can involve more than one NSP or SP. A slice involving more than one NSP raises additional trust challenges and corresponding security requirements [6].

The notion of a domain or a sub-domain is within the jurisdiction of a single NSP or SP.

Multiple sub-domains are plausible within a single NSP or SP. Service categories may be sliced further. The extent to which a service category is sliced is established by the NSP.

A network slice may be composed of virtualised and/or non-virtualised entities.

More than one endpoint device or user equipment may connect to the same network slice (e.g. sensors and infotainment devices/user equipment for automotive).

The 5G system shall allow a common core network associated with one or more access networks to be part of a network slice (e.g. fixed and mobile access within the same network slice).

A Network Slice includes the following:

a) Control Plane functions associated with one or more User Plane functions (e.g. a reusable or common framework of control),
b) Service or service category specific Control Plane and User Plane function pairs (e.g. user specific multimedia application session).

Endpoint or user equipment can connect to a single network slice or to more than one slice. This raises additional security challenges, particularly relating to isolation between slices [6]. The network should control which network slices the user equipment can connect to simultaneously.

When endpoint or user equipment accesses multiple network slices simultaneously, a set of common control plane functions should be utilised by the multiple network slices and their associated resources.

The NSP (e.g. network operator or a virtual network operator, such as MVNO) uses a Network Slice Blueprint to create and manage a Network Slice Instance.

A network slice instance [4] may be;

Wholly statically defined, e.g., as in fixed-access business or residential service, or

Partially dynamic, e.g., as in roaming mobile endpoint/user equipment which may be connected to a statically-defined service chain, or

Fully constructed on demand

Even when a network slice instance is statically defined, the necessary resources may be virtualised e.g., as transport tunnels over a layered infrastructure network, or as VNFs located somewhere in a cloud. The actual physical resources, together with their configuration, may thus vary over the course of time, including on-demand allocation or scaling.

A Network Slice Instance provides the network characteristics which are required by a Service Instance. Examples of a network slice instance include all the three categories of services – eMBB, mIoT, and URLLC – that span human-to-human, human-to-machine and machine-to-machine interfaces, which cover personal, industry, vehicular, social, health, city, and industry services and applications.

Examples of a Sub-network instance could be the IMS (IP Multimedia Subsystem), separate core and access sub-networks provided by different vendors or a subset of network functions within an administrative domain realising parts of the network slice instance.

The Service Instance Layer represents the services (end-user service or business services) which are to be supported.

Each service is represented by a Service Instance.

> NOTE: A Service Instance can either represent an SP or an NSP service. The SP or NSP services may be a 3rd party provided service.

An administrative domain refers to the scope of jurisdiction of a provider. A provider may obtain service capabilities from 3rd parties to enrich the services it provides to its end customers. A provider could also benefit from offering its spare capabilities or resources to a 3rd party.

A network service can be a single user connectivity service, NaaS (Network as a Service) such as a service instance, a network slice instance or a subnetwork slice instance offering for a business vertical that utilises forward-looking business models, or IaaS (Infra-structure as a Service).

The notion of a partnership between two providers is qualified in terms of the one which is hosting the service, and the one whose service is being hosted, as described in the NGMN Network Slicing Concept paper [4]. A formalised description of the roles that qualify the behaviour of a provider is as follows:

Provider-Hosted (P-Hosted): A service provider that provides services to e.g. end customers, which can negotiate with another provider (Provider-Hosting) based on a trust model, for the establishment of a hosted network slice instance or a hosted Sub-network instance using functions and resources from the hosting domain.

> NOTE: The necessary resources, in the hosting domain, are allocated based on a configured SLA between P-Hosted and P-Hosting,

Provider-Hosting (P-Hosting): A service provider, which can negotiate with another provider (P-Hosted) based on a trust model, for providing the usage of functions and resources in the hosting domain towards the hosted domain.

> NOTE: The necessary resources, in the hosting domain, are allocated based on a configured SLA between P-Hosted and P-Hosting,

Different types of partnerships and sharing may be envisioned, with a variety of distinctions:

> Various levels of functional exposure are considered, as envisioned in Section 4.5.2 of the NGMN 5G whitepaper [1].
>
> 5G should provide an abstraction layer as an interface, where all types of in-networking functionality (control plane and user plane related) can be exposed to the application layer functions and/or service providers based on a service level agreement. The application/service provider will then be able to use a sub-set of the network capabilities in a flexible, configurable and programmable manner, and to use network resources depending on their service preference.
>
> Automated real-time negotiations, as well as manual acquisition imply different considerations.

Static or dynamic configuration of a partnership

Partnerships or agreements may be based on one or more bilateral agreements for realising any set of multiple partnerships

A bilateral partnership or agreement is typically based on an SLA (Service Level Agreement) between two parties, where each of the two participating providers are enabled to provide the necessary resources for the realisation of a service instance or a network slice instance.

For scenarios where the services of a broker are leveraged, there would be a pair of bilateral SLAs in place, where the broker behaves as a trusted mediator for the realisation of a service instance or a network slice instance.

5.2 Network Slicing – single administrative domain

The NGMN Network Slicing Concept paper [4] contains the following provisions;

A Network Slice Instance may be utilised by multiple Service Instances provided by the network service provider. This offers economy of scale/less overhead.

NOTE: Whether there is a need to support utilisation of Network Slice Instances across Service Instances provided by different 3rd parties is for discussion in Standards Development Organisations (SDOs).

A Network Slice Instance may be composed of zero, one or more Sub-network Instances, which may be utilised by other Network Slice Instances.

A Sub-network Blueprint is used to create a Sub-network Instance from a set of Network Functions, which run on the physical/logical resources.

A Network Slice Blueprint is used to instantiate a Network Slice Instance.

A “Network Slice Instance-X” may be derived from a composite “Network Slice Blueprint-PQ” that has constituent “Sub-network Blueprint-P” and a “Sub-network Blueprint-Q”.

The “Network Slice Blueprint-PQ” is inherited from the constituents ““Sub-network Blueprint-P” and “Sub-network Blueprint-Q”.

The “Network Slice Blueprint” may also be a simple composition of “Sub-network Blueprints”, where there is no inheritance.

5.3 Network Slicing – multiple administrative domains

The NGMN Network Slicing Concept paper [4] contains the following provisions;

The network slice blueprint may include resources or service capabilities from other providers with which an SLA exists.

In general, there are two categories of scenarios where network services need to be provided across multiple service providers:

Roaming scenario: Individual users move from one provider (i.e. Home NSP), which is the P-Hosted domain to a network managed by another provider (i.e. Visited NSP), which is the P-Hosting domain. The services that a user requires while roaming needs to be specified in the SLA between the two providers. In this case the two providers, with an SLA, would be the P-Hosted domain (Home NSP), and the P-Hosting domain (Visited NSP), with the corresponding behaviours required to support the inbound roamers (e.g. using a service instance or network slice instance) by the P-Hosting domain. The notion of tenancy, where the appropriate functions are provided by a P-Hosted domain to a P-Hosting domain based on SLAs is allowed, to meet the service characteristics required by the P-Hosted domain.

Business verticals: When a business vertical service user's request cannot be met by the capabilities of a single SP, the SP may harness the necessary capabilities from another SP, based on an SLA between the two SPs. In this case the two SPs, with an SLA, would be the P-Hosted domain (Home SP), and the P-Hosting domain (Third-party SP), with the corresponding capabilities required by the P-Hosted domain obtained from the P-Hosting domain.

6 NETWORK LAYER

6.1 Architectural considerations

The 5G system shall support a service-based architecture design, which enables modularised network services. The service-based architecture and interfaces in the 5G control plane make the 5G network flexible, customisable, and independently deployable. NSPs can leverage service-based architecture design in 5G to manage and customise the network capabilities, e.g., by dynamically discovering, adding, and updating network services while preserving performance and backward compatibility (when required). The network services functionality should enable reusability, information hiding, high cohesion, and loose coupling across network services. The service-based protocols should be lightweight.

The 5G core and access networks are to be functionally decoupled to create an access technology agnostic architecture [1]. This preserves the ability for NSPs to source core and access networks from different vendors and maintains core and access network separation, with interoperability provided by standardised interfaces between them. The objective of the 5G architectural framework is to provide the flexibility required to realise the 5G performance targets for different usage scenarios. For example, the reduction in network latency requires the decentralization of computing resources and storage in order to enhance service experience. This implies flexible orchestration of computing and storage resources from centralised to edge/cloudlet infrastructure [45].

The optimal location of computing and storage resources is a trade-off between latency reduction gain and physical security of the assets and stored information. The placement of computing and storage resources far out at the edge of an access network, such as in the case of a split between centralization and distribution (e.g. CU/DU split) achieves the most latency reduction gain, with additional security risks that require to be mitigated, especially if a control plane function is located in a domain, where there is no implicit trust. The placement of computing and storage resources at the edge of the core network instead of the access network sacrifices some latency reduction gain but can be at a physically secure location within the NSP's trust domain. Independent of whether the computing and storage resources utilized are in the core network or in the access network the privacy of user information requires to be ensured.

The tactile internet [5] is a significant area of forward-looking usage scenarios, under the category of ultra-reliable low-latency communication services. A notable requirement for enabling the tactile internet is to place the content and context bearing virtualised infrastructure as close to the user as is securely possible. This direction provides innovative directions for new revenue sharing opportunities and collaborative business models across various flavours of SPs. Content, context, and mobility demands are vital ingredients required to suit the demands of reliability, availability, and low-latency. NGMN has already identified that low latency / edge computing mechanisms also raise additional trust challenges and corresponding security requirements [8]. DLT models that are described in section 6.3, offer potential solutions to some or all of the above security and privacy issues.

The dominant themes within the tactile internet exemplify requirements for reliability and availability that need to be met at the access network edge to suit a variety of services that engage a human-in-the loop, across human-to-machine and machine-to-machine interfaces. The multimedia services in the tactile internet landscape are required to enable haptic interactions with visual feedback that augment the audio-visual user experience. Other multimedia components that are required as relevant for enabling tactile-internet services over a human-machine interface include vestibular and proprioceptive sensory translations. The tactile internet services with these multimedia component augmentations are required to be rendered with imperceptible latency. Such interfaces include robotic and machine-learning systems, with usage scenarios that span industry automation, telepresence, integrative health, autonomous vehicles, education, smart grid, renewable energy, personalisation, entertainment, art, cultural enrichment, etc.

The end-to-end latency required [5] for a satisfactory experience of tactile internet services is in the region of one millisecond, which implies much lower latency contributions over the radio-link segment, and under mobile handover conditions.

Human perception is guided by the sensory apparatus, which provides a measure of the quality of experience of interactions with the surrounding environment. This enables a feedback loop for an adaptation to the environment or to modify the experience of the environment. In the context of the tactile internet a corresponding service is an example of the environment. For a consistent, intuitive, and natural service experience, the service must be adaptable to the response time of human sensory perception.

The requirements to enable these new types of services, with the simultaneous demands of ultra-low-latency, reliability, availability, and mobility present the most challenging class of services that must be supported by an architecture framework that is sufficiently generalised, flexible, scalable, adaptable, and extensible. These architectural considerations are required to be examined at the access and core network layers.

NGMN has published some relevant information relating to the needs of vertical industries in [9].

6.1.1 Consistent User Experience across access networks

The 5G system shall support a consistent service experience across 3GPP and non-3GPP access networks, including in scenarios involving hand off between heterogeneous access technologies. The services may however have to be adapted to access specific characteristics for example in terms of QoS. Operationally a consistent service experience can be facilitated for example by adoption of a common set of procedures and functions for AAA, QoS, Policy, session continuity etc.

User applications should always be connected to a RAT, combination of RATs and/or attachment points (or other user equipment acting as a relay in case of D2D), or combination of points of attachment to the network providing the best user experience without any user intervention, according to NSP/SP subscription and policy.

The 5G system shall be able to provide an Inter-RAT mobility service interruption time that does not degrade the user experience, including between 3GPP and non-3GPP access technologies. The interruption time (if any) as a KPI should be compliant with the expected and desired user experience for the targeted service. For instance, for voice service, the 5G system shall be able to deliver service continuity by ensuring Inter-RAT seamless mobility. Some non-3GPP access networks may provide substantially lower security assurance than 3GPP networks do. Additional assurance may therefore be needed when accessing some core network features or services [8].

6.1.2 Fixed-Mobile Convergence considerations

To enable flexible management and joint optimisation the 5G system requires harmonised fixed and mobile Network Management and Orchestration.

Harmonising different identity and authentication, QoS, policy, and charging paradigms in cellular networks, (wireless) local access networks, and fixed networks is essential to enable the convergence of different access types, to facilitate the realisation of different business models and to maintain a consistent user experience.

An attractive user experience is characterized in terms of convenience and seamless connectivity, across fixed or mobile access facilitated by different technologies (e.g. on the move or at home). Service usage requires an “always best connected” fabric for service delivery, independent of implementation.

Some use cases benefit from Fixed Mobile Convergence (FMC) relative to others. For example, an eMBB service consumed at a mobile device may be served on the move via cellular networks (LTE, 5G NR) and in stationary or indoor environment via Wi-Fi access and fixed networks. For this use case, a network-based system-level traffic steering is advantageous for the user as well as for the NSP. On the other hand, predominantly mobile applications, such as vehicular scenarios (e.g. autonomous cars, connected cars, and assisted driving etc.), would be less relevant from an FMC perspective.

Aspects of FMC with respect to 5G are driven by the target to enable a seamless connectivity for user equipment across different kinds of cellular & Wi-Fi (residential & public) access networks, including the possibility to combine the capabilities of different access networks when available at a certain location. This means, that all use cases that demand a higher bandwidth, higher availability, and higher reliability would inherit multiple access paths for providing ‘better’ connectivity. Equipment comprises both mobile terminal devices and (residential) gateways providing wireless access (e.g. WiFi or Bluetooth) for portables or also wireline (e.g. Ethernet) connectivity to stationary terminals (e.g. TV screen, refrigerator). On the other hand, low data rate best effort services, such as those that may belong to a service category, such as mIoT, may not always be required to leverage FMC for simplicity.

E2E network architecture shall enable a technology agnostic operation by abstracting methods and levels. Interaction with technology-specific or vendor’s specific managed entities is done through encapsulation of configurations, profiles and translation of commands, operations, and notifications.

A network-based approach for connectivity control allows for an optimized resource utilization, and customer experience. Network based control of connectivity shall be applicable, including functionality (e.g. authentication and authorization) and end-to-end traffic steering (e.g. aggregation of traffic, differentiation between different access technologies, selection of different access technologies).This implies that the respective functions will be deployed in the network (e.g. functions for access traffic steering, switching and splitting [36], and that corresponding functionalities will be provided by the user equipment. Hence, a standardized end-to-end solution is required which will be supported by NSPs, vendors, terminal OS and terminal providers.

An FMC ready architecture concept will gain most from modularly assembled Network Services to provide maximum flexibility (i.e. SBA, network slicing, separation of data processing and data storage, one CP and several UPs, etc.) and thus be characterized by the following principles:

- ***Network Service Convergence:*** Use of common 5G services classified as CPS (es) and UPS (es), independent of access technologies.
- ***Optimisation of the User Plane:*** Support for optimization of the user plane to suit the traffic requirements associated with customer services, such as Mobile traffic, Internet traffic, IPTV traffic, Voice, etc., as well as new tactile applications such as Virtual/Augmented Reality (VR/AR). This includes use cases from the industry sector associated URLLC services that demand ultra-low-latency, and high reliability guarantees.
- ***Multi-provider scenarios:*** Use of 3rd party access networks and 3rd party applications and services shall be supported.
- ***Cellular & Wi-Fi Multi-Connectivity on mobile devices:*** Such pre-requisite is seen as most powerful and most flexible approach for implementing network-based traffic steering.

- ***Staged deployment approach*** *(migration path towards target architecture):* Multiple levels of inclusion/integration and co-operation/interworking scenarios shall be supported to fully exploit the advantage of FMC in every stage of deployment (e.g. re-use of existing fixed network technology before 5G-ready fixed technology).

An architectural model of a Fixed-Mobile framework is shown in Fig.3. . In this model a UE is enabled to experience seamlessness while connecting via multiple access networks to multiple UP services.

In principle, the UE can be connected through one or through different access networks at the same time. This depends on the requirements of an end-user service. A UE can be a smart phone, CPE etc. Different connections at the same time might be for different services (e.g. Voice service in parallel with IPTV service and Internet service. Another use case might be a hybrid access approach, where other UPs are set-up, which are not shown in Fig.3. . An example would be where the basic Internet traffic is served by the fixed line and in time of more bandwidth needs, the 5G network will take over the additional requests.

For simplicity, Fig.3. depicts the case, where all UP services for different applications are controlled by the same CP service. However, it is also possible to have a set of CP services that are dedicated for controlling a certain set of UP services. Depending on SLAs and specific customer demands an inclusion of additional CPSs provided by 3rd parties in the framework of collaboration/partnerships may be enabled using service exposure capabilities and correspondingly designed APIs.

A typical example is Authentication and Authorisation (AA). For cellular networks, SIM-based authentication is utilized, and for fixed access networks, a line-ID or/and username and password is utilized. A network service, which is associated with an access network over which the request was initiated, is addressed and the UP is configured appropriately. The Network Services may work together through a common set of data for the customer or will configure the same UP, or will address other common network services, such as for monitoring, QoS configuration, or for an execution of common policies.

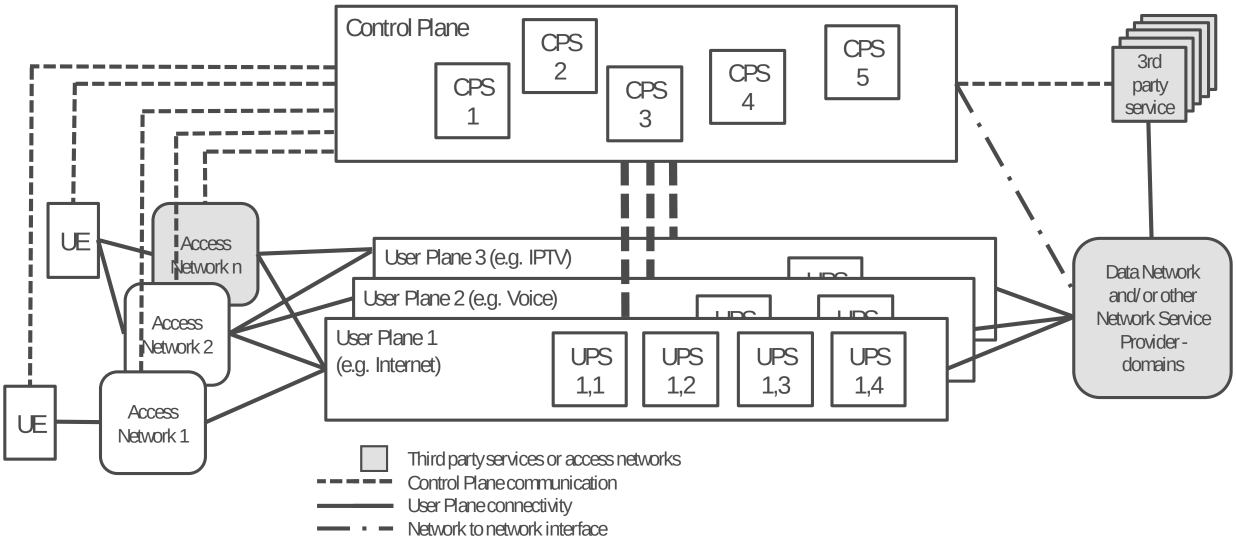

Control Plane
CPS 1
CPS 2
CPS 3
CPS 4
CPS 5
3rd party service
UE
UE
Access Network n
Access Network 2
Access Network 1
User Plane 3 (e.g. IPTV)
User Plane 2 (e.g. Voice)
User Plane 1 (e.g. Internet)
UPS 1,1
UPS 1,2
UPS 1,3
UPS 1,4
Data Network and/or other Network Service Provider-domains
Third party services or access networks
Control Plane communication
User Plane connectivity
Network to network interface

Fig.3.   : Potential architecture for the use of different User Planes controlled by one Control Plane

A concept for such tight integration of multiple access networks as depicted in Fig.3. will demand strong changes to both wireless and wireline technologies and will impact the standardization activities at both 3GPP and BBF for establishing interoperability On the other hand existing deployments which already allow for coexistence of fixed and mobile networks should experience as little impact as possible to avoid any performance impact on current business models. Principles, such as a staged deployment approach in terms of finding the optimum migration path towards an envisioned integrated approach described above are desirable.

6.2 Potential enablers for meeting required Key Quality Indicators

6.2.1 General

There may be several different enablers which could assist in meeting the Key Quality Indicator (KQI) requirements. This section describes some which may need to be developed further by the relevant SDOs.

Capabilities dispersed throughout the end-to-end framework are required to meet diverse KQIs associated with the three main categories of services, namely, eMBB, massive IoT, and URLLC. For example, in the case of the eMBB category of services high data rates at appropriate levels of QoS are a critical requirement, with the associated KQI targets. In case of massive IoT or massive MTC (Machine Type Communications) massive scale, variable payloads of information, low-cost options, battery longevity, low-maintenance, resource constrained operation etc. are among the requirements, with related KPI measures. The combination of high-reliability and low-latency requirements are among the QoS profiles, associated with URLLC services, with the related set of KQIs.

Energy efficiency, virtualisation, transport efficiency, handover efficiency, self-organisation, and enhanced utilisation of resources in the core and radio access networks, are pivotal enabling capabilities.

6.2.1 Minimization or avoidance of tunnel overhead

The core and access networks have traditionally consisted of different entities and sub-networks, where the Internet Protocol (IP) has been simultaneously used as a service locater and a service identifier. With mobility, in the context of a TCP (Transport Control Protocol) session, service interruptions as result of a change in the location of an IP connection is addressed by layer 2, and layer 3 tunneling of packets in-flight.

With the advent of composite access networks (e.g. different radio access technologies) and heterogeneous access networks (e.g. different coverage footprints), the requirement is to provide seamless connectivity in the presence of mobility. This expanding diversity of deployment choices, coupled with ultra-low-latency, reliability, availability, and mobility, demand a reduction in the overhead associated with the frequent setup and teardown of the necessary tunnels in the conventional manner to accomplish seamless service mobility.

The changes in the geographical location of a point of attachment of a device to an access network edge resulting from mobility, would add more overhead with tunneling, in a functionally virtualized network, which would further impair an ultra-low-latency dependent service experience. Hence a minimization of tunneling overhead or the avoidance of tunneling overhead are required to scale as needed, while satisfying the most stringent requirements associated with tactile internet services.

A separation of the control plane and the user plane combined with the notion of using common tunnels for similar types of services minimizes the control plane overhead. Dynamic instantiation of virtualized control plane and user plane functions allows different levels of centralization and distribution to meet assorted service experience demands.

The avoidance or minimization of tunneling reduces the control plane signaling overhead, in terms of required resource utilization, related fault potential, mobility, and latency, in a virtualized environment [6]. The separation of the location identity of device, which is dependent on the topology of an access network to which the device is attached, from the domain dependent identifier for the device, provides a mechanism to avoid the tunneling overhead. The domain dependent identifier for the device is access network topology independent. This implies that while the device may change its topological attachment point, its domain dependent identifier is unchanged. The mapping of the domain dependent identifier of a device to a location identity is a function of mobility, or other criteria that may result in a change the location identity. [7].

The guiding principle here is that the ‘name’ or ‘identifier’ of a resource, such as a device, indicates ‘what’ is sought, while an ‘address’ indicates topologically ‘where’ it is, and a ‘route’ indicates ‘how’ packets arrive and depart from the device. The separation of an 'identity' from the 'address' provides an approach to remove the overhead of tunneling.

A separation of the location or address at the point of attachment of user equipment to a network, which is dependent on network topology, from a topology invariant name or identifier of the user equipment is useful to avoid the overhead of tunnel setup and management. This would be especially relevant for ultra-low-latency services in the context of edge computing and mmWave high-bandwidth, small footprint coverage arrangements.

Along the thinking of a separation between the location and a topology invariant identifier for user equipment, the objective of the Identifier Locator Separation (ID-LOC) architecture is to simplify the management of network, devices, and sessions by employing two namespaces: Identifier for the user equipment's identity and Locator for its location in the network [38].

Another area for tunnel optimization is the IoT use cases involving the transfer of small amounts of data on an infrequent basis. In many such scenarios the IoT devices involved are also stationary. An example of that would be utility meters that send out a meter reading of a few hundred bytes on a monthly or configurable time interval basis. Not only do the meters not require tunnelling for mobility support, the signalling overhead for establishing a tunnel with a subsequent tearing down of the session after transmission of few hundred bytes introduces overhead and reduces the system and spectrum utilization efficiency. For such use cases a connectionless data transfer approach involving a shared tunnel that is established for carrying data for all such devices from a base station to the core network is a beneficial choice.

A dedicated user plane connection with a tunnel for each device would be avoided for such use cases. The device can request such a connection type at the time of registration with the network and afterwards no signalling is required during any data transfer over the related UP. With this approach, since there is no signalling, the RAN does not get involved in data security aspects and they are negotiated directly between the device and the core network. This approach is part of an ongoing study pertaining to 5G IoT [39].

Further exploration of security aspects with respect to tunnel avoidance or minimization for enhanced system efficiency is required for examining the trade-offs. For example, from a security perspective the network would be required to interpret the headers created by the UE (e.g. for routing, which would expose a new attack surface, requiring new mitigation strategies). The impacts, in terms of the flexibility of new functionality to accommodate the transport protocols, mobility and charging, are for further study.

6.2.2 Optimisations for Edge Computing and Fixed/Nomadic uses

Edge Computing and Nomadic / Fixed access (wireless or wired) are two of the key 5G usage scenarios. Optimisation of the 5G framework for such cases must therefore be considered. For example, in the case of edge computing, the changes in the geographical location of a point of attachment of endpoint/user equipment to an access network edge resulting from mobility would add more overhead with tunnelling in a functionally virtualised network, which would impair an ultra-low-latency dependent service experience. Hence a minimisation of tunnelling overhead or the avoidance of tunnelling overhead may be required to scale as needed, while satisfying the most stringent requirements associated with tactile internet services. Similarly, when the endpoint/user equipment is stationary in its lifetime such as in the case of fixed 5G wireless deployment, paging it and use of tunnelling for the sake of mobility support can be viewed as an unnecessary overhead and complexity.

6.2.3 Microservices

The notion of microservices for applications is a significant enabling concept in the end-to-end-framework. As the name implies, a microservice is a small autonomous service that has its own architecture, technology, and platform. This type of service lends itself to distributed realisation of applications. The service can be managed, deployed and scaled in an independent manner throughout its lifecycle. For example, a microservices enabler could be applied for a realisation of desired functionality or customisation associated with customer experience, data analytics etc. The service can also be constructed from other combinations of building-block applications [10].

The benefit of utilising the microservices concept is that it is an enabler to suit various types of business models and contexts, in a manner that complements other enabling facets of the end-to-end framework, such as virtualisation and edge computing. Further details and applicability of these concepts are for further study.

6.3 Distributed Ledger Technology (DLT)

6.3.1 General

The notion of a distributed ledger in DLT consists of a secure database for tracking physical or virtual resources, without requiring a centralized mediation. In other words, data is securely accessed or exchanged across distributed entities, such as different administrative domains in a peer-to-peer manner. Disintermediation of a centralized entity or a third-party broker, for any transaction of value between any two entities, or among a defined group of entities that share a DL, is an intrinsic characteristic of a DL capability.

The benefit of a distributed ledger is that it has the potential for minimizing the latencies and the costs associated with data access or exchange. This capability also enables entities for automatically utilizing smart contracts, without intermediaries by providing near real-time evidence of any tampering, which in turn provides a framework for compliance with relevant policies or a regulatory regime.

6.3.2 DLT concept

A distributed ledger leverages the blockchain framework. The contiguous records stored in a distributed ledger are divided into blocks, where each block is chained to the next block using a cryptographic signature. Each block of data may contain one or more records, which are part of the distributed ledger. A record block can only be added, when the participants of the distributed ledger establish a consensus in terms of the validity of a record block. The way the blockchain framework establishes a consensus, on any proposed change in the distributed ledger, defines the type of the blockchain that is utilized.

Each record in the distributed ledger is date and time stamped, together with an associated unique cryptographic signature. This unique cryptographic signature ensures the authenticity and the integrity of the distributed ledger. The protection of each record in the distributed ledger obviates the need for any central or trusted intermediary to process the records in the distributed ledger for validating the records and related transactions, in terms of authenticity and integrity.

6.3.2.1 Smart Contract

As part of a distributed ledger framework, a smart contract facilitates, verifies, and automatically negotiates agreements between any two entities that share a distributed ledger. The transactions associated with a smart contract are traceable and irreversible.

6.3.2.2 Blockchain Types

The type of blockchain that is utilized is based on the process for establishing consensus pertaining to any proposed addition of records information in a distributed ledger. The process may be either permissioned or permission-less. For example, some distributed ledgers may be permission-less implying public access, while others may be permissioned to only allow private and restricted access to individuals or entities for portions of the distributed ledger or the entire distributed ledger. The distributed ledger must be designed in a manner that enables a rapid detection of anomalies associated with data changes in the distributed ledger because of unauthorized access or a fault condition.

Some of the use cases for the application of a distributed ledger span interdisciplinary service categories, such as self-executing smart contracts, digital health insurance, clearing and settlement in payment systems, property registration, financial services, micro-insurance, credit provisioning etc. While some of the use cases are more forward-looking than others, some of the existing use cases such as the establishment and execution of SLAs (Service Level Agreements) to suit roaming scenarios across different NSPs, or SPs, can be examined in terms of smart contracts that leverage the use of a distributed ledger.

The distributed ledger with its inherent blockchain framework is an emerging concept in the NSP or SP space, from the point of view of telecommunications. Within the E2E framework, an E2E 5G service may span different types of blockchains that constitute an associated distributed ledger. There is an absence of standards in the industry with respect to the distributed ledger concept, in terms of interoperability. The use cases in the industry that utilize the blockchain have been published [40]. A proof-of concept initiative termed as: "Unleashed Blockchain", has been created in the industry, based on the use-cases in [40] that include the following five significant usage scenarios:

- Roaming
- Identity Management
- SLA (Service Level Agreement)
- Mobile Number Portability
- Stolen mobile devices

These use cases provide guidance for delineating the related requirements from an E2E perspective, in terms of the distributed ledger concept and the blockchain framework, in the context of the network slicing building block.

The use cases listed earlier are within the scope of the E2E 5G Architecture and exploring them in the light of the 5G Slicing would help the 5G community.

Further studies in terms of trade-offs and benefits associated with the distributed ledger concept are anticipated [37].

6.3.2.3 DLT considerations

The DLT capabilities that leverage the blockchain framework shall comply with conformance and certification methods delineated by SDOs for SPs to adopt interoperable behaviours for accelerating the adoption of blockchain oriented products and solutions to suit 5G services within the E2E architectural framework. [Add reference – IEEE etc.]

Test specifications for establishing inter-operability and system integration should be developed in the industry for usage by SPs. This would enable SPs to deploy distributed services, over multi-vendor and multi-layer blockchain platforms, while benefitting from incremental innovation in the longer term. Furthermore, implementation specific service differentiation and hence value creation, over an interoperable DLT capabilities, would be an attractive direction for SPs.

The criteria for selecting candidate use cases, the industry shall specify common and shared criteria to serve as a guide for identifying an adoption of DLT capabilities [50] [51].

The SDOs shall establish assessment procedures on the application of DLT capabilities for use cases associated with 5G services within the E2E framework.

Automated procedures should be specified to convert any existing business logic or legal contracts into smart contracts, within DLT capabilities.

Synergy and complementarity between DLT capabilities and AMC (Autonomic Management and Control) should be leveraged to serve the following objectives:

- Cognitive business
- AMC for adaptability and programmability of resources and services
- Decentralized trust, traceability, transparency, privacy, and security
  Privacy should be combined with the attributes of trust, traceability, transparency, and security in a customizable manner, where DLT, AMC, and artificial intelligence can be leveraged to meet common objectives [51].

6.4 Autonomic Networking (AuN)

6.4.1 General

The essence of a 5G ecosystem consists of a diversity of network technologies and deployment arrangements that are required to support a plethora of services, which have a vast variety of quality of service, quality of experience demands, while assorted content and media types are exchanged with the user equipment from an E2E perspective. This fundamental aspect of a 5G ecosystem, underscores an indispensable need to adopt a framework of autonomic capabilities, from an E2E perspectives, for an automating the configuration, awareness, and operation of 5G networks and systems.

The relatively static, best-effort, or case-by-case integration and operation of networks in the previous vintages of mobile or fixed wireless communication systems are inadequate for meeting the diverse and dynamic, provisioning, and operational complexities of the 5G ecosystem. The conflicting requirements of extreme flexibility, dynamic adaptation, and enhanced resource utilization [1] to suit diverse service demands require optimization through a framework of autonomic capabilities.

An autonomic function is inherently adaptable on its own, or in other words autonomously to a changing environment. The self-managing attributes of an autonomic function are characterized in terms of auto-discovery of information and entities that it requires to drive self-* features such as self-configuration, self-protection, self-healing, and self-optimization of network resources, services and parameters, which are the MEs that are managed and controlled by the autonomic function [41] [42] [46]. The process of automation is enhanced through the use of an autonomic framework, where decision-making does not require human intervention.

A generic autonomic networking architectural model, such as that described in [41] [42] is a hierarchical model based on two nested and complementary control-loop levels from the perspective of time scale and objective. It is a generic model in the sense that it defines and separates generic concepts and associated architectural principles in terms of autonomic networking, cognitive networking and self-management that encompass methods, implementation strategies, and related details, suitable for flexible deployment scenarios.

In an autonomic framework, a higher-level entity such as the KP (Knowledge Plane) [42] provides system-level awareness across disparate subsystems or domains. The definition, concept, behavior, and the reference points associated with the KP have been specified [42].

Closed-loop control is a fundamental attribute of an autonomic framework, where disparate subsystems or domains are capable of interacting in a peer-to-peer manner, which implies that conventional top-down schemes or protocols are incompatible to serve the needs of such spontaneous interactions that include a level of dynamism in terms of discovery, as needed. Discovery is a typical attribute of autonomic functions.

Note: Additional considerations and security aspects will be further studied and delineated in a future version of this document.

6.4.2 Use cases for an Autonomic Management and Control (AMC) framework

The network slicing enabler serves as a foundational building-block in the 5G system, which demands compatible strategies for managing resource utilization and performance in an automated and autonomous manner to optimize operational efficiency, service assurance, and the user experience [1]. The instantiation of network slices, for diverse services and a variety of usage scenarios in a scalable manner, with specific run-time constraints, including with an E2E scope, demands a management framework with autonomic capabilities. As an example, ITU-T “Requirements and Architectural Framework for Autonomic Management and Control of IMT-2020 Networks” [53] adopts and recommends a use case to realize the AMC architecture through a generic autonomic networking architecture reference model Architecture [42] reference model.

The flexibility demands in a 5G ecosystem, requires a separation of ownership and the capability to manage and control specific network resources through policies, such as in the case of federated domains or other domain partitions. These capabilities were not considered in the design of earlier vintages of cellular communication systems. A dynamic management and control subsystem, autonomic capabilities with cognitive awareness is indispensable to meet the flexibility and adaptability attributes, which are pivotal for rendering diverse 5G services, over a virtualized, decentralized, and distributed 5G deployment scenarios.

The inherent diversity of 5G usage scenarios with their related constraints, in terms of latency, pace of mobility, mixed media types, radio resource demands etc., autonomic frameworks provide a various levels of context awareness in terms of connectivity and resource allocation for optimized service rendering behaviors, for a 5G service-based system architecture, which must also be localized to enhance the end-user service experience, through MEC arrangements.

Among a virtually unlimited range of usage scenario stories, some of the distinguishing attributes of diverse and dynamic service scenarios include the following:

- Interactive and complex user-centric services (real-time, and non-real-time)
- Security and privacy
- Mobility context
- Predictability (e.g. network capacity, and coverage)
- Federated domain collaboration and cooperation
- Network performance, and resource utilization efficiency
- Service quality of experience from an end-user perspective
- System energy efficiency (e.g. network and user-equipment)
- Machine-learning and analytics
- SLA enforcement
- Cognitive and autonomous tuning of network and service optimization
- Ease of deployment, scaling, optimization, and migration in a system enabled by network slicing and virtual functions for a rapid time-to-market

Some illustrative examples of usage scenarios for autonomic management and control suitable for consideration in a network sliced 5G E2E framework are exemplified in [52].

6.4.3 Autonomic Management and Control (AMC) concepts

AMC [41] [42] is about Decision-making-Elements (DEs) as autonomic functions (i.e. control-loops) with cognition introduced in the management plane as well as in the control plane, whether these planes are distributed or centralized.

Cognition (learning and reasoning used for advanced adaptation) in DEs, enhances DE logic and enables DEs to manage and handle a variety of scenarios, including unforeseen situations and events detected in the environment around the DE(s).

Control refers to the control-logic, which forms the kernel of the DE that realizes a control-loop to dynamically adapt network resources and parameters or services in response to changes in network goals/policies, context changes and challenges in the network environment that affect service availability, reliability and quality. The DE contains this type of control -logic.

DEs realize self-* features (self-configuration, self-optimization, etc.) because of the decision-making behaviour of a DE that performs dynamic/adaptive management and control of its associated Managed Entities (MEs) and their configurable and controllable parameters. Such a DE can be embedded in a network node (Network Element (NE) in general) or at a higher layer of the outer overall network and services management and control architecture. An NE may be physical or virtualized (such as in the case of the NFV paradigm).

Autonomic management is distinct relative to automated management. The former emphasizes learning, reasoning, and adaptation, while the latter focuses on efficient workflow implementation and automation of the processes involved in the creation of network configuration and monitoring tasks.

Automated management provides input to the AMC. Indeed, AMC must exhibit a network governance interface through which the input that governs the configuration of an autonomic network should be provided. AMC exposes views and reports to the automated management process.

Fig.4 illustrates the positioning of both paradigms and highlights the interaction between them [41].

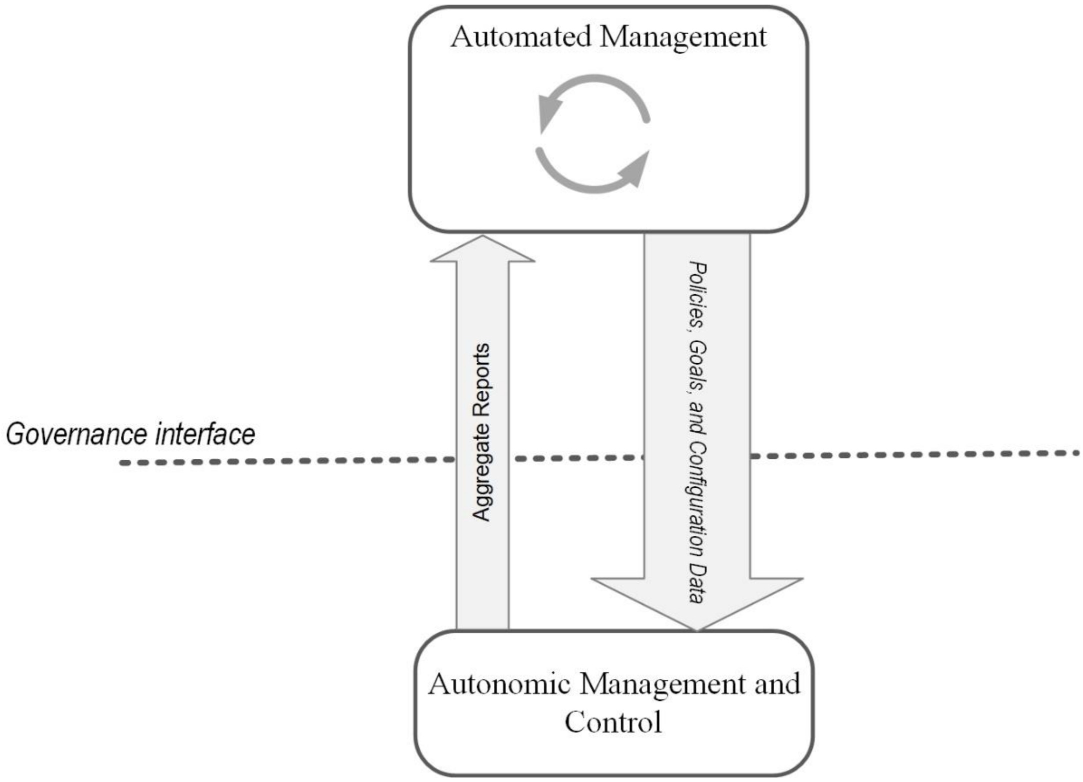

Automated Management
Aggregate Reports
Policies, Goals, and Configuration Data
Governance interface
Autonomic Management and Control

Fig.4. Automated Management vs Autonomic Management illustration (their interaction and complementarity)

The concept of AuN includes self-manageability and properties within network nodes/functions and "in-network" self-management. It includes Autonomic Management and Control (AMC) as described in [41] [42].

An exemplification of the entities engaged in a cooperative framework of automated management and an autonomic networking environment, illustrating the notion of a Knowledge Plane (KP) is depicted in Fig.5.

To achieve this objective, there is a need for a standardized Federated Framework to support E2E Autonomic (Closed-Loop) Service Assurance as specified in [41] [42].

The guiding principles for an autonomic networking framework for delineating the related requirements include:

- The interworking of management in the context of a hybrid network that consists of an AD and a non-AD associated with a legacy OSS shall be supported.
- The network analytics functions performed by interworking modularized Autonomic Managers (AMs) using the and associated shared Knowledge Bases (KBs), which execute as software with real-time and predictive analytics as loadable modules or applications, shall be supported.
- Multi-Layer autonomics, such as the abstraction levels at which autonomic and cognitive capabilities can be embedded, shall be supported. For example, this multi-layer architecture includes E2E Service Orchestrator, SDN controllers, OSS, NFVO, VNFM, VIM, EM, virtualization layer as a multi-tenant platform

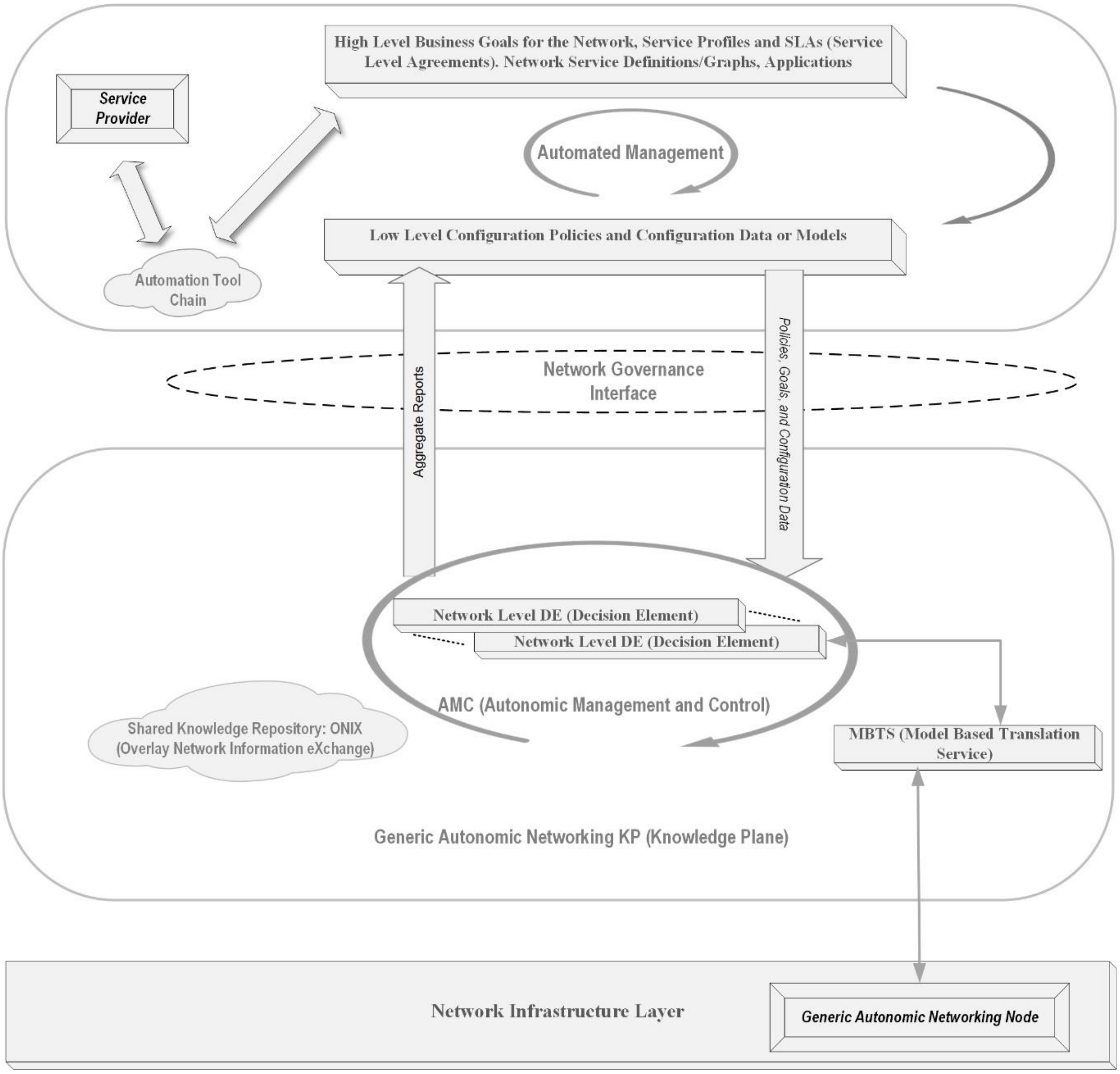

High Level Business Goals for the Network, Service Profiles and SLAs (Service Level Agreements). Network Service Definitions/Graphs, Applications
Service Provider
Automated Management
Low Level Configuration Policies and Configuration Data or Models
Automation Tool Chain
Network Governance Interface
Aggregate Reports
Policies, Goals, and Configuration Data
Network Level DE (Decision Element)
Network Level DE (Decision Element)
AMC (Autonomic Management and Control)
Shared Knowledge Repository: ONIX (Overlay Network Information eXchange)
MBTS (Model Based Translation Service)
Generic Autonomic Networking KP (Knowledge Plane)
Network Infrastructure Layer
Generic Autonomic Networking Node

Fig.5. Illustration of entities in an autonomic management and control environment

This E2E autonomic service assurance of E2E network services shall be achievable through a federation of specific network segments/domains access, backhaul, and core networks all embedding autonomic and cognitive capabilities. The scope of a federation may be extended to cover other Autonomic Domains (ADs), where such domains have embedded autonomic capabilities beyond the core network, such as a data center network that hosts some Telco-Cloud network functions or even applications.

6.4.3.1 Fast and slow control loop concept in AuN

A hybrid AuN framework is required, to suit an autonomic management of a diverse array of networks and services that characterize a 5G ecosystem. This type of framework for autonomic management of networks and services, its hybrid nature is characterized in terms of hierarchical levels of abstraction, consisting of ‘fast control loops’ and ‘slow control loops’.

Fast control loops are localized control loops that require a fast reaction time (e.g. control-loops that operate within a Network Element/Function (NE/NF), in the case of the management of entities within or closer to a radio node or user with strict 5G latency constraints etc.).

Slow control loops are broad scope control loops that require a network-wide view, such as in the case of joint management of entities further away from a radio node or user with relaxed latency bounds that are typically associated with non-session scenarios etc.

A time-scaling of control loops provides a generalized capability to suit a wide range of feedback control reaction times in an AuN framework. The main objective of ‘slow control loops’, consists of a management and control subsystem process that extends beyond the network and service entities, with respect to long-term goals for the network and service behavior, characterized in terms of network and service optimization, adaptation and the dynamic planning of a service or resource (re)-orchestration in to respond automatically to various challenges experienced by the network, including changing workloads dynamics, manifestations of faults/errors/failures and performance degradations). [48][49] Insights into management and control aspects [48] [49]are addressed through fast control-loops to operate within NE/NF and those that should be addressed by slow control-loops that have a broad scope should be considered in operational aspects for long-term planning and policing of network and service behaviours.

Autonomic feedback control-loop based service assurance for network services, including for network slicing, has profound benefits in terms of 5G network performance optimization, quality of experience, and operational efficiency

The design and interworking of ‘fast control loops’ and ‘slow control loops’, where the latter layers of abstraction hierarchically overlay the former layers, form a conceptual framework for multilayer autonomics, as specified in a generic autonomic networking architecture model[42] that describes such interworking, where the higher-level autonomic control-loops (slow control-loops) with wider views of global target objectives perform “policy-control”, while the fast control-loops are implemented in NEs/NFs. The related requirements form the basis for a holistic, interoperable AuN framework.

6.4.3.2 Hierarchical abstractions in AuN

The control loops or self-management constructs in AuN are abstracted in terms of multiple hierarchical levels in a holistic manner for realizing the concept of multilayer autonomics, which enables the following capabilities:

- Guidance on an instantiation of an autonomic network model on various network architectures and their associated management and control architectures (e.g. [48] [49]).
- Guidance on trust and confidence in autonomics by addressing the stability of control loops and the quality of the decision-making logic and algorithms of embedded autonomic functions
- Guidance for DE implementations in terms of the expected variation in the complexity of AI (Artificial Intelligence) oriented cognitive algorithms, together with room for further improvements in the DE algorithms over time, depending on the DE’s hierarchical level of operation.

6.4.3.3 Decision Element (DE)

Autonomic behaviours of a DE include a secure auto-discovery of the following items:

- Network objectives and policies specified by the human operator

- Other DEs it requires to collaborate with in terms of ownership
- Capabilities of the DE's assigned Managed Entities (MEs), such as the information that is available at run-time.

After auto-discovery, a DE performs the self-* operations on its assigned MEs as arranged by design, by orchestrating (launching and/or configuring) the MEs when required, and adaptively (re-)programming the MEs as required via their management interfaces.

Orchestration implies the launching (at run-time) of an entity (e.g. an ME) if no instance exists to provide a desired service, via a configuration of the service instances (the newly launched or an already existing one) such that the entity is ready to provide a service that the entity is designed to provide.

A DE is designed to perform one or multiple self-* operations such as self-configuration, self-diagnosing, self-healing, self-repair, self-optimization, self-protection, etc. Some specialized DEs may be designed to perform certain self-* operations on a macroscopic level that considers wider perspectives needed to complement the same self-* operations intrinsically performed by DEs on the microscopic level (e.g. VNF, PNF, 5G SBA NF, 5G SBA Service etc.).

It serves as a blueprint model that defines and prescribes the design and operational principles of autonomic Decision-making Elements. It is similar to the Hybrid-SON model because it shares common principles with the Hybrid SON architectural model, as they both enable combining and interworking centralized and distributed management and control solutions for networks and services.

As an example, a Hybrid SON framework adopted for some deployments is compatible with the autonomic architectural principles. The Hybrid SON [47] being compatible with the autonomic architectural model is described in [41] [42] model. This autonomic architectural model has been instantiated over various network architectures and their associated management and control architectures to enable DE implementation enhancements for a realization of autonomic capabilities in terms of control loops. Examples of a generic autonomic architectural model [41] [42] include:

- Autonomicity and Self-Management in the Broadband Forum (BBF) Architectures (an instantiation of generic autonomic networking architecture model in BBF architecture scenarios) [48].
- Autonomicity and Self-Management in the Backhaul and Core network parts of the 3GPP Architecture (a GANA instantiation onto Backhaul and EPC Core Network) [49].

6.4.3.3.1 Triggers for DE operation

The triggers for a DE operate in collaboration with other DEs whenever required, are various input information and changes that drive its algorithms (e.g. machine learning, deep learning, computational science and other types of decision-making algorithms or AI algorithms).

Such input information can be changes in the operational state of its ME(s), changes in the governing input policies, context changes, and challenges (e.g. faults, errors, failures) detected in the operation of the MEs and the underlying network substrate.

The collective, and collaborative autonomic operation of the DEs on orchestrating and adaptively programming (configuring) the various MEs in the network to achieve a global network objective, with respect to the desired instantaneous operational state of the MEs, is a realization of the notion of self-organization of the required management and control operations.

6.4.3.4 AM (Autonomic Manager)

The DE is also referred to as an Autonomic Manager element [41]. The autonomic manager element is a functional entity that drives a control-loop meant to configure and adapt (i.e. regulate) the behaviour or state of an ME (i.e. a resource).

6.4.3.5 ME (Managed Entity)

The ME refers to a managed resource, as distinct from a managed element, which is a term used in a traditional network management context normally intended to mean only a physical Network Element (NE) and not some functional entity within a node/device such as a protocol module, or mechanism. An NE may be physical or virtualized, such as in the case of the NFV paradigm.

6.4.3.6 CM (Cognition Module)

A DE that exhibits AI (Artificial Intelligence) algorithm(s) for autonomics is referred to as a "Cognitive DE". Regarding cognition property, Two options are provided [42] regarding whether to embed a Cognition Module (CM) for Data Analytics in a DE (i.e. a DE having its own embedded CM), or to have a CM being commonly shared among some DEs.

Multiple DEs (e.g. KP DEs or DEs introduced in a Network Function or in a service in the SBA case) may have a commonly shared Cognition Module (CM) that implements Data Analytics, such that the Cognition Module's analytics outputs are used by the DEs in their decision-making processes for autonomic operations on programming (re-configuring) their associated MEs. Therefore, a CM as a Data Analytics Module, implements cognitive algorithms that operate on raw data to synthesize knowledge that can be represented in certain formats and streamed to DEs (if it is a commonly shared CM and not a CM embedded in a DE) to aid the DE(s)' autonomic operations and decision-making logics.

6.4.3.7 Autonomics algorithms

Autonomics algorithms are meant to be implemented by the DEs [48] , and such algorithms include cognitive algorithms for Artificial Intelligence (AI), such as Machine Learning (ML) and Deep Learning (DPL), together with other algorithms that can be employed in to facilitate the closed-loop operations of a DE.

6.4.3.8 Mechanism as a functional entity

A "mechanism" as a functional entity within a node/device is to be understood in general as task execution logic, that can be invoked/launched to perform a sequence of tasks that the mechanism is designed to perform. When launched, it has a lifecycle involving a start-state of its execution and usually a final termination-state of its execution upon the completion of the tasks or jobs it is designed to accomplish, whenever it is launched and allowed to execute to its completion.

A mechanism is designed in such a way that it can be orchestrated (launched). It can be implemented as an executable software program or software library function(s) that can take various forms, such as for example a protocol, TCP/IP application layer or OSI layer 7 application layer entities. However, as noted in [42] there are mechanisms that in general cannot be classified as protocols or TCP/IP application layer or OSI application layer entities.

In a dynamic management of network services and parameters, various types of mechanisms can be dynamically orchestrated as Managed Entities (MEs) by AM components and employed to achieve some objectives. Those kinds of mechanisms are ones that are manageable, via an exposure of management interfaces.

Examples of such managed mechanisms in autonomic management and control of network services and parameters include the following:

- IP Packet Capturing Mechanism
- IP Packet Classification Mechanism
- Fault Detection Mechanism
- Fault Isolation/Localization/Diagnosis Mechanism
- Fault Removal Mechanisms
- Active and Passive Probing mechanisms
- Link failure detection mechanism
- Automated module replacement mechanism.

6.4.3.9 Autonomic behaviours of DE

Autonomic behaviours of a DE include a secure auto-discovery of the following items:

- Network objectives and policies specified by the human operator
- Other DEs it requires to collaborate with in terms of ownership
- Capabilities of the DE's assigned Managed Entities (MEs), such as the information that is available at run-time.

After auto-discovery, a DE performs the self-* operations on its assigned MEs as arranged by design, by orchestrating (launching and/or configuring) the MEs when required, and adaptively re-)programming the MEs as required via their management interfaces.

Orchestration implies the launching (at run-time) of an entity (e.g. an ME) if no instance exists to provide a desired service, via a configuration of the service instances (the newly launched or an already existing one) such that the entity is ready to provide a service that the entity is designed to provide.

A DE is designed to perform one or multiple self-* operations such as self-configuration, self-diagnosing, self-healing, self-repair, self-optimization, self-protection, etc. Some specialized DEs may be designed to perform certain self-* operations on a macroscopic level that considers wider perspectives needed to complement the same self-* operations intrinsically performed by DEs on the microscopic level (e.g. VNF, PNF, 5G SBA NF, 5G SBA Service etc.).

It serves as a blueprint model that defines and prescribes the design and operational principles of autonomic Decision-making Elements. It is similar to the hybrid-SON model in terms of sharing common principles with the hybrid SON architectural model, where both enable combining and interworking centralized and distributed management and control solutions for networks and services. As such, the hybrid SON (now being deployed) is compatible with the design principles inherent in the generic autonomic networking architecture model [41] [42]. Details[47] pertaining to a hybrid SON are compatible with the generic autonomic networking architecture model [42].This model has been instantiated in various network architectures and their associated management and control systems to enable DE innovators to implement the prescribed DEs that enable a realization of autonomics (control-loops). Examples of generic autonomic networking architecture realizations based on the principles identified in [42], include:

- Autonomicity and Self-Management in the Broadband Forum (BBF) Architectures that reflect a realization of [48], in a BBF architectural scenario.
- Autonomicity and Self-Management in the Backhaul and Core network parts of the 3GPP Architecture that reflect a realization of [49], for backhaul and EPC (Evolved Packet Core) network architectural scenario.

6.4.4 Federated E2E decision making across autonomic domains

The responsibility demarcation or scope of business and technical and technical aspects, within the E2E federated architecture are associated with the infrastructure or network slice level (see section 11) or at the orchestration level in terms of inter-domain and intra-domain scenarios (see section 10).

The demarcation and collaboration across multiple 5G domains is enabled through federated arrangement of autonomic domains. This facilitates a static or a dynamic demarcation based on policy and design principles for a conflict-free autonomic architecture. [42]

6.4.4.1 Considerations in autonomic domains

The requirements for autonomic domains are delineated in terms of the following:

- Requirements pertaining to the design principle of AuFs autonomic functions and decision making in the E2E architectural framework
- Requirements pertaining to information associated with identification and KPIs that are exchanged at federation reference points

6.4.4.2 Design principles for AuFs (Autonomic Functions)

The availability of a standardized framework that enables E2E autonomic (closed loop) service assurance for services, including network slices is foundational for SPs. The E2E autonomic service assurance spans the service design phase and the service execution phase.

The requirements are delineated in terms of the design principles for AuFs and an associated holistic hybrid model for the interworking and complementarity of multi-layer AMCs (Autonomic Management and Control) functions at various levels of abstraction in a network management and control architecture.

6.4.4.3 E2E decision making process and governance

A minimization or avoidance of human intervention or mediation in terms of management and control is beneficial for optimizing operational efficiency for SPs. The following is a list of requirements necessary for E2E autonomic service assurance of E2E network services, including 5G network slices, through a federation of KPs (Knowledge Planes) specific to network segments and domains.

E2E Autonomic (Closed-Loop) service assurance should be achievable through a federation of Knowledge Planes (KPs) that implement components for AMC intelligence for specific network segments (viewed as domains). Autonomics rendered by the KP in a generic autonomic networking architectural model is intended to be complemented by lower level autonomics introduced in Network Functions (NFs) associated with any particular network segment/domain within the scope of the KP, such that the KP controls the policies that are relevant for the lower level autonomics introduced in NFs.

The E2E federations of KPs for the various network segments/domains and their policy-controlling of lower level autonomics (fast control-loops) in the NFs associated with the respective network segments, enables complementary multi-layer autonomics. Such an arrangement allows for a realization of holistic multi-domain state correlation and resources programming by the KPs for the associated network segments/domains, such as for access, X-Haul (i.e. Fronthaul, Midhaul and Backhaul), core networks, etc.

The low-level DEs injected to operate in CP functions (or CP services in the case of 3GPP SBA) and UP functions (or UP services in the case of fragmented UPF) of a network segment or domain should be policy-controlled by their “mirror-DEs” in the KP responsible for the network segment/domain (e.g. a “QoS Management-DE” operating in a UP function or CP function should be policy-controlled by a mirror “QoS Management DE” in the KP level),

Therefore, the KP applies to both CP and UP, as it needs visibility of events and state concerning both CP and UP. [47] [48] [49] provide illustrations of the relevant principles for a realization of a federation of KPs based on the principles for federated autonomic management and control prescribed in [42].

6.4.4.4 Information exchanged at federated reference points

A common or generic information set and KPIs for exchange across reference points or domain boundaries is useful for SPs or NSPs, for any use case under consideration. This generic information set and KPIs require a common and shared specification, in terms of the associated primitives and procedures. This approach mitigates anomalies, simplifies the design processes, while facilitating performance assessments and benchmarking.

In addition to a common and shared specification, it is anticipated that the former would be complemented by specific requirements for instantiating a generic federated AMC framework for application in existing network architectures, and their associated management and control architectures.

In a federated arrangement of DEs, a DE-to-DE interface may be positioned at different points in the network topology and at the borders between network segments or domains. Such an interface across DEs would apply in implementations within the KP [42].

The peer DEs discover the following types of information associated with their peers at the reference points pertaining to a federated AMC [42]. The following categories of information are exchanged at a federated AMC reference point:

- ➢ Requirements for common and generic information at the federated AMC reference point.
  - ▪ This is associated with the identification of peer domains engaged in a federation of domains that embed autonomic and cognitive capabilities.
- ➢ Requirements for KPIs to be exchanged at the federated AMC reference point.
- ➢ Requirements associated with a federated AMC reference point.

6.4.5 AI (Artificial Intelligence) and ML (Machine Learning) in AMC

The significance of AMC in forward-looking E2E system is in its various offering of self-CHOP (self-configuring, self-healing, self-optimizing, and self-protecting) characteristics combined with a variety of flexible and self-adapting cognitive capabilities, for automatically adapting to a changing environment. These qualities of the AMC framework enable the simultaneous satisfaction of diverse business model and system performance objectives in an optimized manner.

ML provides an algorithmic approach for AMC to learn, adapt, and improve continuously based on feedback control loop constructs [54]. ML is a dominant approach for embedding human intelligence characteristics, within the broad field of AI through applications of computers and computing processes.

The feedback control loop constructs within AMC are integrated with shared knowledge information, which is enmeshed with other building block entities within the KP, which is a part of AMC. ML is a pivotal technology within AMC, where the E2E system behaviour is adjusted in response to a perception of the E2E system environment and the processed information, experience, and intelligence within the KP. Adaptation, based on this cognitive process underpins the characteristics of the KP, which enables the system to automatically select different and appropriate functionalities, by leveraging different software functions. This type of behaviour cannot be realized through statically provisioned rules (e.g. software-based rules), since a continuous prediction of a changing system environment is both complex and infeasible.

The KP processes input data from a variety of sources in the E2E system to infer relevant and current knowledge through continuous learning, prediction, and clustering models in ML [55]. In ML, clustering models group data into clusters, based on similar characteristics, for deriving inferences within the KP.

6.4.5.1 Stakeholders and Multilayer AMC with cognitive capabilities

The cognitive capabilities embedded within a DE utilize AI and ML models of learning and responding to dynamic shifts in the system and environmental contexts, through slow and fast feedback control loops. The DEs are part of the KP in the AMC framework with various relationships and interfaces associated with a variety of human stakeholders that represent diverse roles and functional responsibilities. The consumer of DEs could interact with vendors/suppliers for changes or updates, via a related catalog or marketplace of advertised deployable AI and ML models, utilized within a DE.

The design and execution of cognitive capabilities within AMC are performed via the relationships and interfaces associated with the AMC as illustrated in Fig.6.

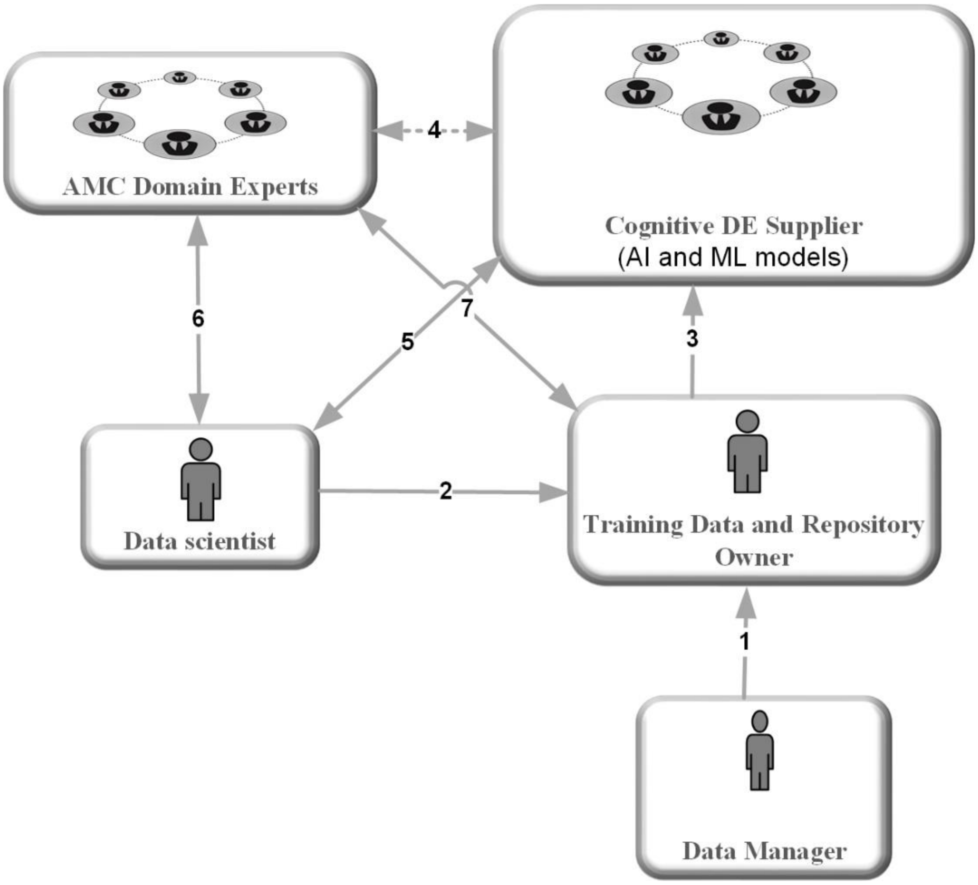


Fig.6. Stakeholder interfaces and relationships with a cognitive AMC framework

The human stakeholder engagement and oversight with AMC framework is minimized and delineated by the specific stakeholder roles. The arrangement illustrated in Fig.6. derived from [57] is such that infrequent or no intervention required, where an infrequent intervention is limited to changes in business objectives or updates to the functions in the autonomous framework, as a result of continuing technology advances. An exemplification of a sequence of stakeholder interactions, associated with the AMC, is also shown in Fig.6. . The sequence of numbers and arrows in Fig.6. is illustrative and aligned with the corresponding stakeholder roles to exemplify the principles associated with the nature of the subject and the related roles and interactions.

The role of each stakeholder enumerated in Fig.6. is summarized below:

- **Data Scientist**

The Data Scientist requires the competencies associated with an interpretation of the value of data, through the use of a repository of cognitive algorithms that utilize an iterative process of data ingestion, knowledge synthesis, and clean training data for customer delivery.

- **AMC Domain Experts**
  The AMC Domain Experts engage the Data Scientist to harness the various aspects of knowledge extraction, and training data from raw data for use in the development of AI (Artificial Intelligence) and ML (Machine Learning) models, and for certain autonomic capabilities that may be required by cognitive DEs.

- **Cognitive DE Supplier**
  The Cognitive DE supplier collaborates with the Data Scientist to get support in preparing training data required to train AI and ML models that are new or evolved models, such as the cognitive GANA DEs for AMC, where a Cognitive DE supplier may not have their own Data Scientist to perform that task.

  The Cognitive DE supplier collaborates with the AMC Domain Experts, where the Cognitive DE Supplier does not have their own AMC Domain Experts for the design of autonomic management and control algorithms and control loop logic.

- **Training Data and Repository Owner**
  The Training data and Repository Owner can be played by an organization that is trustable and neutral, such as standards bodies or industry forums (e.g. TM Forum, ETSI, BBF, ITU-T, 3GPP, GSMA, NGMN, IEEE, etc.) or even Open Source Projects.

  The input data for training cognitive DEs based on AI and ML models are retrieved from the training data repository.

- **Data Manager**
  The Data Manager collects and formats raw data and manages data, while interacting with a training data repository owner and the repository. The Data Manager may be any trusted entity that has a trust and business relationship with the training data repository owner.

6.4.6 AMC KP (Knowledge Plane) use cases

The structure and constituents of AMC accommodate federated arrangements of network domains, whether these domains are sub-domains within a single service provider or whether they are domains associated with different service providers. The coordination and distribution of AMC framework behaviours across a federation of domains require multilateral policy and service level agreements across the participants in the federation, such as in the context of network slicing, where a network slice may be hosted, where the hosting domain is required to meet the requirements in terms of performance and service related KPIs associated with the hosted domain.

The autonomic entities within AMC provide the capabilities for an automation of management and orchestration in a service-based 5G system, where network slicing is a pivotal enabler for end-to-end service rendering and isolation for an effective realization of security and fault management. These autonomic entities, consisting of different types of DEs manage a variety of software/hardware elements, using feedback control loops at different levels of knowledge abstraction, with functional characteristics that embody monitoring, analysis, planning, and execution. These capabilities that are intrinsic to AMC allow for an automated instantiation of network slices for diverse end-user service scenarios in a scalable and autonomous manner, with specific run-time constraints, while fulfilling an end-to-end scope, quality of experience KPIs, and business imperatives.

Automation through AMC is indispensable for effectively managing complexity and changing system environment conditions, resulting from increasing levels of feature sophistication, function interdependence, network heterogeneity, and edge computing, with end-to-end scope in a 5G system.

The self-CHOP type of characteristics intrinsic to AMC are reflective of the cognitive capabilities enabled through the various modalities of ML, where the E2E framework of 5G is provided with the ability to adapt based on both historic information as well environmental change pertaining the 5G system. This implies that AMC is capable of optimizing the system behaviours based on relevant metrics to establish an optimal configuration and performance, while recovering and healing from faults to avoid system outages.

The KP (Knowledge Plane) within AMC is representative of a cognitive network management system, based on AI and ML constructs. With the movement of intelligence and services to the edges of a network infrastructure, coupled with a need for increasing levels of distributed and heterogeneous networks to provide experiential advancements to service quality in the 5G ecosystem, the role of the KP is pivotal. The KP resolves the inherent deficiencies associated with the simplistic nature of the end-to-end principle, which assumes that there is no need for the core of a network infrastructure to be aware of when and where any information is created or consumed, such as at the edges of the network. The use of the KP in AMC obviates this assumption and enables network arrangements for strategically managing complexity and scale required to support system optimization, architectural flexibility, and end-user service quality in a heterogeneous 5G infrastructure.

The principles embodied in the KP of an AMC framework consist of the following:

- A separate plane, relative to the control plane and the user plane
- Harmonization of the flexible and diverse needs of the system, such as resource allocation for different services
- Cognitive capabilities through machine-learning and artificial intelligence for decision-making in cases with limited information, or conflicting information

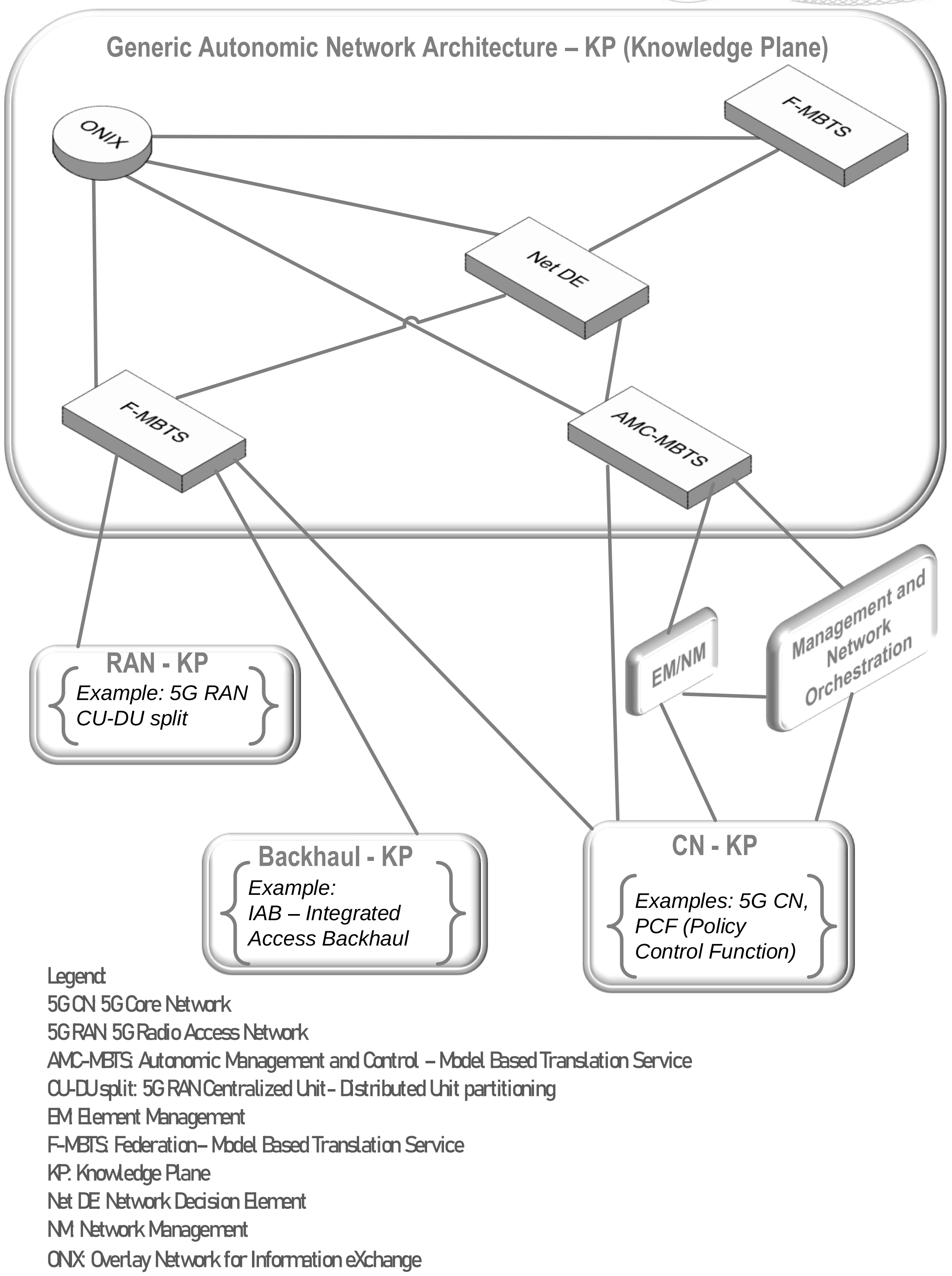
AMC is complemented the aspect of network orchestration, which regulates the allocation of network resources based on a variety of management decisions. In concert with AMC, orchestration is required to support a variety of AMC use cases. Examples of a few AMC use cases are:

- E2E Autonomic (Closed-Loop) Service Assurance for 5G Network Slices by Cross-Domain Federated AMC Knowledge Planes
- Autonomic Monitoring
- Autonomic Security Management and Control
- AMC in automated and intelligent management and control operations of Broadband
- Autonomic Management and Control of routing and energy saving

6.4.6.1 AMC KP use case for mobile access

The principles of a generic autonomic networking architecture are relevant in a forward-looking mobile access system [49], from an E2E perspective for a realization of self-CHOP capabilities required to manage and evolve a complex and diverse ecosystem of services, while optimizing system performance, resource utilization, and end-user service experience. The instantiation of these principles in 5G and beyond systems are suitable for application in the core network, access network, and backhaul segments of an E2E architectural framework that adopts AMC realizations in variety of arrangements with a single domain or in federation of multiple domains. Federated generic autonomic networking architecture KPs provide the knowledge plane capabilities and configurations for each of the domains within a federated arrangement together with the network slicing enabler [56].

An exemplification of an AMC architectural framework in a 5G mobile access system is illustrated in Fig.7.


Generic Autonomic Network Architecture – KP (Knowledge Plane)
ONIX
F-MBTS
Net DE
F-MBTS
AMC-MBTS
EM/NM
Management and Network Orchestration
RAN - KP
Example: 5G RAN CU-DU split
Backhaul - KP
Example: IAB – Integrated Access Backhaul
CN - KP
Examples: 5G CN, PCF (Policy Control Function)

Legend:
5G CN 5G Core Network
5G RAN 5G Radio Access Network
AMC-MBTS: Autonomic Management and Control – Model Based Translation Service
CU-DU split: 5G RAN Centralized Unit - Distributed Unit partitioning
EM Element Management
F-MBTS: Federation – Model Based Translation Service
KP: Knowledge Plane
Net DE: Network Decision Element
NM Network Management
ONIX: Overlay Network for Information eXchange

Fig.7. Exemplification of AMC KP in a 5G mobile access system

6.5 Access Networks

The 5G core network will support multiple access networks including both fixed and mobile. FMC (Fixed-Mobile Convergence) is considered important (covered by requirements in all the following sections). Additionally, the 5G system will support the use of non-3GPP access for off-loading and maintaining service continuity. The 5G network shall enable the placement of applications taking latency or relevance to a defined geographical area into account.

Multiple connectivity (e.g. through multiple access technologies, or different links associated with the same access technology), where available, shall be supported to optimise resource allocation and signalling.

Customer convenience should be a major aspect in providing access technology independent or converged services. As such service specific demands shall be analysed by an access overarching entity taking the decision on an overall efficient selection of the (near) optimum composition of the transport path.

Different access technologies complement each other in some scenarios – e.g. fixed access would provide off-loading to the scarce mobile network capacity e.g. in in-door environment. On the other hand, a flexible allocation of mobile resources could enhance limited fixed access connectivity temporarily (off-loading from fixed to mobile access). The latter variant is used in “Hybrid Access” [34] for Home Gateways (i.e. bundling of fixed access & cellular access). This enables a dynamic distribution and leveraging of capacity across fixed and mobile connectivity.

6.5.1 Mobile Access Network

The 5G system will allow multiple Radio Access Technologies (RATs) to be deployed and enable the seamless introduction of new RATs along with the flexible management and joint optimisation of radio frequency resources [8]. The redundant duplication of RAN functions for different RATs should be avoided, potentially through the unification of common RAN functions for different RATs. The simultaneous utilisation of multiple RATs by system users should also be enabled.

The 5G system will also support flexible RAN structures including implementations based on Cloud principles and the placement of context awareness closer to the RAN edges (i.e. multi-access edge computing). Both centralised and distributed implementation of RAN functions should be enabled to facilitate the realisation of various RAN implementations. In addition, support for various coverage layers and cell sizes spanning extreme long-distance covering macro cells to small cell radio access deployments is required.

6.5.1.1 RAN Decomposition

Functional decomposition of the radio network is required to meet the diverse information transport demands (high performance to low performance) and align them with the demands of next-generation service categories of eMBB, mIoT, URLLC and other new arising categories. To accommodate them a decomposition of the radio network protocol layer functions, across layer-1, layer-2, and layer-3 is required, in terms of the degree of centralisation or distribution [12].

This decomposition consists of placing more functions of the upper layers of the radio network protocol stack in distributed entities for high performance transport demands (e.g. high bandwidth, high-capacity, low-latency, low-jitter etc.) relative to a centralised entity. Scheduling optimisation at a centralised entity, for high performance transport across multiple distributed entities (e.g. base stations, remote radio heads etc.) for fast coordination is a critical requirement.

For relatively low performance transport, more of the upper layer of the radio network protocol stack is placed at a centralised entity to optimise the cost/performance trade-off, associated with the distributed entities. This choice of functional split will determine the X-Haul capacity requirement and associated latency specifications and performance. This will impact the network architecture as it could determine the placement of nodes and the distance between them. A higher layer split will be tolerant of a large latency from a RAN perspective, which may be excessive when low-latency services are considered. Therefore, bounds must be applied within the network architecture to enable a service provider to support low latency services.

A distributed RAN (D-RAN) with several functional splits will be supported by 5G. Fig.8. illustrates the configuration with co-located centralised unit (CU) and distributed unit (DU). All radio protocol layers are terminated within the cell site.

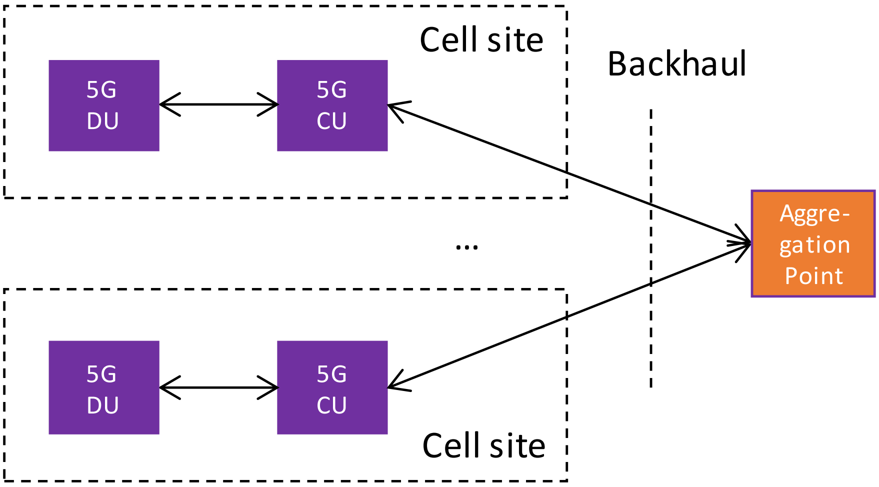


Fig.8. D-RAN Configuration within a cell-site and across cell-sites

The connection from the cell site towards the core network (e.g. via transport aggregation points) is traditional mobile backhaul which will be scaled and optimised to support 5G data rates and performance targets such as low-latency, low Packet Error Loss Rate (PELR), low and very deterministic Packet Delay Variation (PDV) etc. The D-RAN configuration does not constrain the ability of the local CU to support remote DU; in fact, the cell site could become a CU for other cells sub-tended as illustrated in Fig.9. .

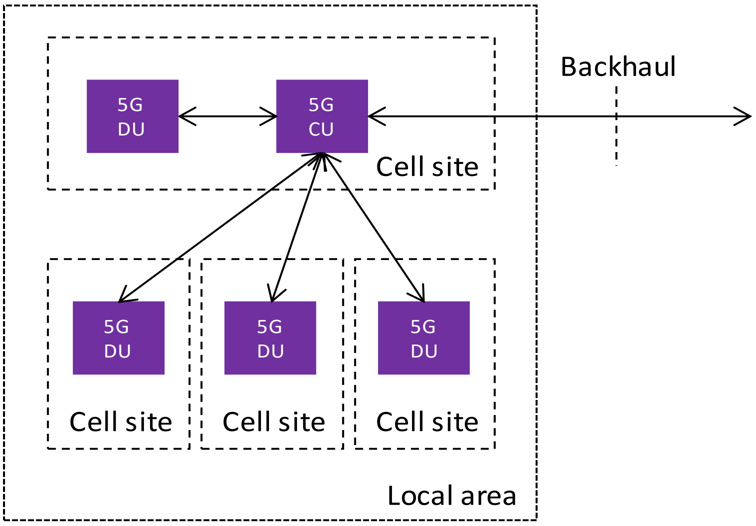


Fig.9. D-RAN with sub-tended DU forming a local 5G C-RAN cluster with shared CU

A 5G C-RAN can be implemented with a higher layer split with protocol stack functions of Packet Data Convergence Protocol (PDCP) and above (e.g., Radio Resource Control (RRC) in the Control Plane and Service Data Adaptation Protocol (SDAP) in the User Plane) being in the CU while the remainder of the stack is in the DU, as shown in Fig.10. . This is one example; other splits will result in a different distribution of functionality between CU and DU.

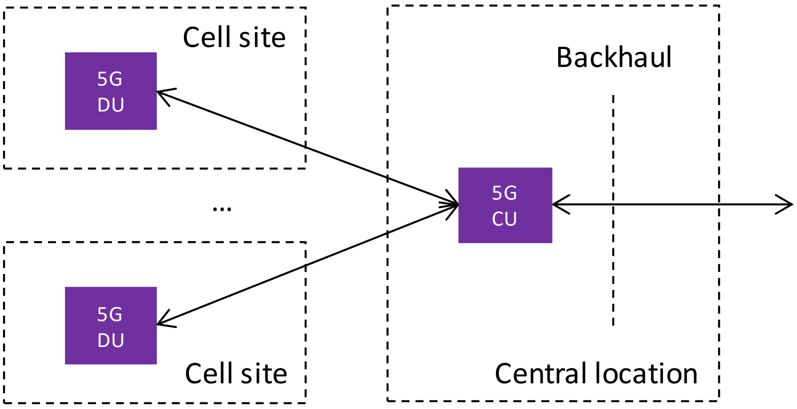


Fig.10. C-RAN functional split

This configuration has similar X-Haul capacity requirements when compared with traditional backhaul. The latency and performance requirements of the RAN are not stringent and therefore consideration must be given to engineer the X-Haul link in accordance with service-based latency and performance targets.

6.5.2 Fixed Broadband Access Network

When applied to Fixed Broadband use cases the 5G system will provide provisions to improve user Quality of Experience (QoE) and maximise the efficiency of service delivery. Examples include Customer Premise Equipment (CPE) with higher capabilities than user equipment, reduced signalling to take advantage of the static placement of CPE and higher performing radio access configurations to exploit channel characteristics under static/outdoor CPE placement.

6.5.3 Wi-Fi Access Network

Among non-3GPP access technologies to be supported by 5G RAN is the 802.11 family, including current 802.11 releases (e. g. 802.11 ac and 802.11 ad) along with future releases (e. g. 802.11ax and 802.11ay). The 5G system shall provide provisions that ensure seamless access point integration, user access and mobility/flow management for Wi-Fi access technologies. This implies a need for automatic/SON-like solutions for fixed access management and orchestration.

As outlined in section 6.1.2 the enhanced user perception experiencing better connectivity by multiple paths in parallel is a 5G feature denoted by FMC. Both trusted and untrusted non-3GPP access should be under consideration in the long term to widen the footprint as much as possible.

6.5.4 Small Cells

The 5G system will enable the seamless integration of small cells under various deployment (such as planned NSP deployment and autonomous deployment) using wired or wireless backhaul. Autonomous deployment of small cells implies a need for automatic/SON-like solutions in small cell management and orchestration.

Small cells in the 5G system should be provided with effective interference cancellation means to enable their operation in the same frequency bands utilised by overlaying macro cells (i.e. co-channel interference) along with other bands not utilised by overlaying macro cells.

6.6 Core Network

The core network in the 5G system shall allow a user to access a network service, independent of the type of access technology.

6.6.1 General

The 5G core network will support multiple access networks including both fixed and mobile types of access networks and thus support FMC as described in section 6.1.2.

The 5G system will provide termination points or points of attachment in the core, for both control plane and user plane information. These points are selected based on location, mobility, and service requirements. They may dynamically change during the lifetime of a service flow. To achieve a converged core network, common mechanisms of attachment should be supported for both 3GPP and non-3GPP access networks.

The 5G system will allow simultaneous multiple points of attachment to be selected for endpoint/user equipment, on a per-service flow basis.

The 5G system shall include a mechanism which provides network discovery and selection based on user experience, reliability and availability demands associated with the requested service.

6.6.2 Control and User Plane separation

Control and User Plane functions should be clearly separated with appropriate open interfaces defined among these types of functions.

6.6.3 NSP controlled Packet Data Networks

In 5G NSP controlled Packet Data Networks (PDNs) are connected to the User Plane Function (UPF) in 3GPP Core Networks via the N6 interface. Such PDNs may be virtualised and SDN controlled. In those scenarios the PDN functions (e.g. content filters, video optimisers, firewalls, DDoS protection etc.) shall be available as VNFs.

6.6.3.1 Decentralisation of Core Network Functions/Core Network Function Services

For reasons of improved latency, storage, or content delivery performance the 5G system shall support placement of core network functions and core network function services and/or storage in physically secure locations intermediate between the centralised core network and the physical locations of access network functions. Such locations are identified as Aggregation Points by ETSI ISG MEC [45] and should be inside the SP or NSP's security trust domain. As such the edge of the core network can be redefined as the boundary between these intermediate locations and the access networks.

6.6.4 Fragmentation of User Plane Function (UPF) Services

Extending the service concept to the user plane enhances the 5G core network in terms of high flexibility, efficiency and programmability. There are several benefits for introducing UPF services:

- ❖ Support cloud friendly deployments of UPF services with finer granularities and independent modules, which can help take advantage of cloud-native services, programmability, and flexible deployment. The service-based framework can be enhanced to facilitate UPF services management (e.g. service registration, discovery etc.)

- ❖ Help customize user plane processing flexibly. There are multiple UPF functionalities (e.g. MPTCP, DPI) that have been investigated and designed in 3GPP, and these functionalities should be deployed on-demand. UPF service design can help with dynamic and flexible deployment arrangements of these functionalities, through user plane service chaining.
- ❖ Efficiency improvement of service communication and ease of service capability exposure. An introduction of UPF services in the service-based framework facilitates direct communications between UPF services and other CPF services, which can prevent duplicate data transfers and reduce transmission paths. This in turn can also help with the ease of original status retrieval or reduce the latency for real-time service flow information from the UPF. Furthermore, UPF services, such as event exposure services, can be introduced to help with UPF capability exposure.

7 BUSINESS ENABLEMENT LAYER

7.1 General

The business enablement layer is a library of all functions required within a converged network in the form of modular architecture building blocks, including functions realised by software modules that can be retrieved from the library for use at a desired location, with an appropriate set of configuration parameters required for certain parts of the network, e.g., radio access. Fig.11. from [1] is illustrative of the context.

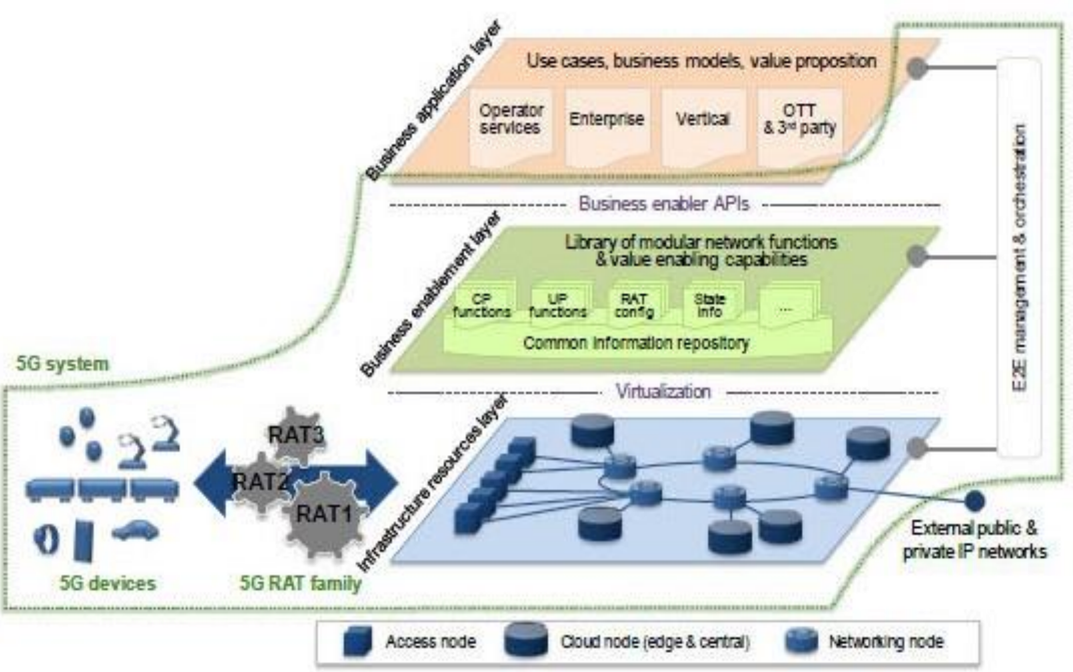

Business application layer
Use cases, business models, value proposition
Operator services
Enterprise
Vertical
OTT & 3rd party
Business enabler APIs
Business enablement layer
Library of modular network functions & value enabling capabilities
CP functions
UP functions
RAT config
State info
...
Common information repository
Virtualization
Infrastructure resources layer
E2E management & orchestration
5G system
RAT3
RAT2
RAT1
5G devices
5G RAT family
External public & private IP networks
Access node
Cloud node (edge & central)
Networking node

Fig.11. 5G system context

The functions and capabilities are called upon request by the orchestration entity, through relevant APIs. For certain functions multiple variants might exist, e.g., different implementations of the same functionality which have different performance or characteristics.

Business Enablement Layer functions shall be realised as virtualised network functions (VNFs) according to the principles specified by ETSI ISG NFV. Their lifecycle management and orchestration shall also be as described there, and their virtualisation requirements documented and managed according to the ISG NFV's VNF Descriptor (VNFD).

3GPP VNFs shall be implemented as user plane/control plane specific 5G entities developed in alignment with Control/User Plane Separation.

The whole architecture facilitates faster introduction of new applications and services compared to traditional networks based on monolithic elements. The separation inherent in its layered structure still allows the use of the most appropriate industry best practices (i.e. for security or content management) at each layer.

7.2 Control Plane Functions

In an end-to-end multi-access network, the control plane functions are not limited to the 3GPP control plane. Control plane functions from fixed, WiFi and 5G mobile access and the converged 5G core network (including non-3GPP functions) are all in scope,

7.3 User Plane Functions

User plane functions include those from fixed, 5G mobile and WiFi access networks, the converged 5G core and NSP controlled PDNs connected to it, implemented as standalone user plane VNFs.

7.4 Configuration Data

Configuration data for each VNF is managed as per the procedures specified by ETSI ISG NFV. For 3GPP 5G functions it is managed as per the procedures jointly agreed between ISG NFV and 3GPP SA5. For non-3GPP functions equivalent processes analogous to the 3GPP versions are required.

8 BUSINESS APPLICATION LAYER

8.1 General

The business application layer contains the specific application packages and services of the NSP, enterprise, verticals or third parties that utilise the 5G network. In virtualised environments, it can be hosted in datacentres or on a Multi-access Edge Computing (MEC) host.

The interface to the management and orchestration system informs the Management and Orchestration (MANO) system of the required composition of dedicated network slices for an application, or the mapping of an application or service to existing network slices. The NGMN white paper on 5G [1] specifically left the detailed contents of this layer out of scope, as can be seen from Figure 1 of that document.

However, the interface to the end-to-end Management and Orchestration system is in scope. Management and orchestration for the application layer is required in a manner analogous to that for the Business Enablement Layer. The implication is that application and service layer software can be orchestrated and managed just like VNFs. Application and service software must therefore inform the Orchestration and Management system of its own infrastructure and runtime environment requirements just as a VNF would via the VNF Descriptor.
Accordingly, network slices can be created, orchestrated, and managed which contain all the physical and virtual network functions and application software required to deliver an end-to-end, multi-layer service.

It is desirable that applications conform to a standard, industry best practice, API format to ease their instantiation and to engage the widest possible community of application developers.

8.2 NSP Applications

NSP applications provide regular telecommunications services such as voice, messaging and internet access, as well as the NSPs own differentiating services which are offered to its own subscribers. The 5G system must include a mechanism whereby NSPs can rapidly instantiate, upgrade and remove new applications and new versions of existing applications, to trial new services and expedite upgrades or rollouts.

8.3 Enterprise Service Applications

NSPs offer service hosting to their enterprise customers. The 5G system must include mechanisms for enterprise service application packages, authenticated and authorised by the NSP, to be instantiated into the business application layer. From there they can form part of a bespoke end-to-end multi-layer enterprise service. A relevant example is business-critical applications, which might reside on premise in the enterprise and/or on a public/private 3rd-party cloud. In such scenarios, the enterprise can own and control its own application packages and the business application layer (e.g. Augmented Reality (AR) maintenance application, secure conferencing/collaboration application hosting etc.). Mechanisms to enable such scenarios must be included in the 5G system.

8.4 Vertical Service Applications

Some 5G use cases are realised by standalone private networks managed by the vertical industry itself rather than the NSP. A good example is factory automation. In such scenarios, the vertical can own and control its own application packages and business application layer. The 5G system must include mechanisms to enable this.

8.5 Authorised OTT and 3rd Party Service Applications

The 5G system shall include support for NSPs which offer service hosting for authorised 3rd party and OTT applications. The host can be a datacentre or MEC host. The 5G system must include mechanisms, by which the OTT player or 3rd party can request instantiation of, and management and usage reports from, their own applications.

9 END-TO-END MANAGEMENT AND ORCHESTRATION

9.1 General

NGMN's publication on 5G Network and Service Management including Orchestration [13] contains many specific requirements on the Management and Orchestration system. They cover such areas as Optimisation, Slice Management, Automation, and Self-organising functionality amongst many others.

The wide range of quality of service demands implicit in a 5G ecosystem demands a corresponding suite of enabling attributes that consist of interoperability, flexibility, extensibility, agility, and dynamism in the allocation of resources, within or across domains that leverage a virtualised environment.

9.2 Orchestration environment

The realisation of a wide range of quality of service demands over a virtualised environment requires the establishment of a network slice that utilises the appropriate resources necessary to support any given service realisation. A network slice may be established within an administrative domain (NSP or SP), or it may be established across multiple administrative domains (NSP or SP).

Service Instances [4] are managed and orchestrated at the Service Instance level.

All the constituent resources of a network slice instance are required to be visible to a 5G management and network orchestration sub-system, including non-virtualised resources (e.g. antenna elements, other elements at any layer of the protocol stack.) For a network slice that includes non-5G functions co-ordination between existing O&M systems and the 5G network management and orchestration system enables migration and coexistence strategies for a coherent and end-to-end management of a network slice.

The virtualisation of specific parts of functions, associated resources and elasticity should be managed at VNF/NFVI level by 5G management and orchestration implementations which are aligned with ETSI NFV MANO standards, to maximise interoperability. For each VNF application FCAPS management shall comply with the FCAPS management specifications published by the SDO which developed it, where possible.

Application and service layer software shall be orchestrated and managed in an analogous fashion to VNFs.

Application and service software shall inform the 5G management and orchestration system of its own infrastructure and runtime environment requirements just as a VNF would via the VNF Descriptor.

The 5G system shall include automatic/SON management and orchestration solutions for the deployment of non-3GPP technologies such as WiFi and their access points.

The 5G system shall include automatic/SON management and orchestration solutions for the autonomous deployment of small cells.

The 5G system shall be capable of managing and orchestrating network slice instances at the granularity of a network, sub-network or service. The interfaces for managing and orchestrating network slice instances and their constituent functions shall be open and standardised to enable interoperability of management and orchestration for the network functions and network slice instances.

9.3 E2E automation

The 5G system distinguishes itself from 4G and previous generations by the pervasive availability of automation for all facets of network deployment and operation. This type of automation is referred to as zero-touch management of network components as well as the network services that run over them in an end-to-end fashion.

The 5G system shall include all operational processes and tasks (e.g., planning and design, delivery, deployment, provisioning, monitoring and optimisation) executed automatically, ideally with 100% automation and without human intervention. This automation becomes an integral part of operators' future Operation and Support Systems (OSS).

The 5G system shall support full automation in a horizontal end-to-end context which refers to across administrative domains such as next generation radio access, 5G core network, and backhaul/transport networks, as well as a vertical end-to-end context which refers to across layers and technologies such as NFV, SDN, and customer facing services.

The 5G system shall support closed feedback loops for life cycle management operations of network components based on performance and fault monitoring, as well as, operator policies both within each domain and across domains to realise an end-to-end 5G service. Such feedback loops shall operate in a fully automated manner with standardised interfaces involving different vendors' network components. Root causes analysis of faults and identification of performance trends, optimisation and proactive orchestration measures to avoid future issues shall be supported with current rule based and emerging Artificial Intelligence based approaches.

The 5G system shall support the fully automated deployment and instantiation of network components to realise an end-to-end 5G system including radio access, 5G core network, and backhaul/transport networks.

Automation requirements also apply to network slices. The 5G system shall support full automation of network slice life cycle operations such as scale-in/scale-out, and update of slice configuration and provisioning of network slices. Furthermore, the 5G shall have the capabilities to monitor the status and performance of network slices as part of an automated feedback loop to maintain network slice performance without human intervention.

Future-proof, horizontal and vertical end-to-end operable framework enabling agile, efficient and qualitative management and automation of emerging and future networks and services are being envisioned [43]. Horizontal end-to-end refers to cross-domain, cross-technology aspects. Vertical end-to-end refers to cross-layer aspects, from the resource-oriented up to the customer-oriented layers. The goal is to have all operational processes and tasks (e.g., delivery, deployment, configuration, assurance, and optimization) executed automatically, ideally with 100% automation.

The objective is to facilitate coordination and cooperation among relevant standardization bodies and open source projects. The Open Network Automation Platform (ONAP) is developing an open source implementation of a platform for real-time, policy-driven orchestration and automation of physical and virtual network functions. Its aim is to be able to rapidly automate new services and support complete lifecycle management [44].

10 ORCHESTRATION ARCHITECTURE FLAVOURS

The following architecture flavours are considered.

10.1 Vertical (Hierarchical) Orchestrator collaboration: layering view

Orchestration shall be multi-layer (vertical/hierarchical) in nature as it involves processes that start from the business level and inductively trigger lower level resource instantiations where synchronisation, delegation or escalation between orchestration layers may be needed. One possibility is that the actions of an orchestrator in one layer may also need to be synchronised with a higher-level orchestrator or for delegation / escalation purpose.

10.2 Federated Orchestration

When considering slices that are provisioned over multiple operators' networks or over multiple domains (sub-networks) within a single operator's network, an assumption of a single top-level orchestrator that has end to end visibility and control over all the domains and networks may not necessarily be true. This is more prominent across different operator/administrative domains, while in scenarios where the service is provisioned across technology domains operated by a single operator - hierarchical orchestration is more likely to be considered as an option. To construct such multiple domain service in the absence of a top-level orchestrator, the individual domain orchestrators must be federated in a manner that allows them to interface with each other horizontally for propagating slice policy and enforcing related rules. It is not necessarily involving hop-by-hop orchestrators along the orchestration path. This may imply some level of coordination / cooperation of autonomic decision-making aspect attached to orchestrators (Intent-based). In an environment where different domains may be operated using different controllers/orchestrators, the use of an industry-wide harmonised Information Model and industry wide standardised east-west-bound APIs is imperative.

10.3 Hybrid Federated and Hierarchical Orchestration.

Actual deployments may include a mix of federated and hierarchical orchestration where certain parts of the end-to-end service are orchestrated by a centralised orchestrator that controls the lower layers vertically, while such centralised orchestrators interface with their neighbour orchestrators horizontally in a federated manner.
Clearly the expectation is that regardless of the underlying method of orchestration, be it federated, hierarchical or a mix of both, the end user should receive ubiquitous experience, no matter how many operators may be involved in the delivery of service and the orchestration methods and approaches used.

11 NETWORK SLICE DEPLOYMENT MODELS

11.1 Categories of administrative domain configurations

The main categories of administrative domain configurations, for the establishment of a network slice, consist of the following:

a) Intra-domain
b) Inter-domain
c) Multi-domain

Each of these configurations can participate in the establishment of a network slice.

An intra-domain configuration refers to one or more sub administrative domains that are provisioned, within a single administrative domain to suit domain specific policies to handle different types of services. A network slice can be established to support a service, within a sub administrative domain. If a service requires the support of multiple sub administrative domains, then the network slice, required to support the service, is established through a cooperation of one or more designated orchestrators, based on the polices associated with the single administrative domain.

An inter-domain configuration refers to two different administrative domains that are required to cooperate to provide the necessary resources and functions to support any given service. The network slice required to support the service is established through a cooperation of the domain specific orchestrators, based on policies and agreements that are applicable across the two different participating administrative domains.

A multi-domain configuration refers to more than two different administrative domains that are required to cooperate to provide the necessary resources and functions to support any given service. The network slice required to support the service is established through a cooperation of the domain specific orchestrators, based on policies and agreements that are applicable across all the different participating administrative domains.

For each of these categories of administrative domain configurations, the participants in the establishment of a network slice would include a combination of SPs and NSPs, or just NSPs, depending on the nature of a given service realisation.

11.2 Network slice arrangements

A contextual view of the main categories of administrative domain configurations, for the allocation of the required resources for the establishment of a network slice, is shown in Fig.12. .

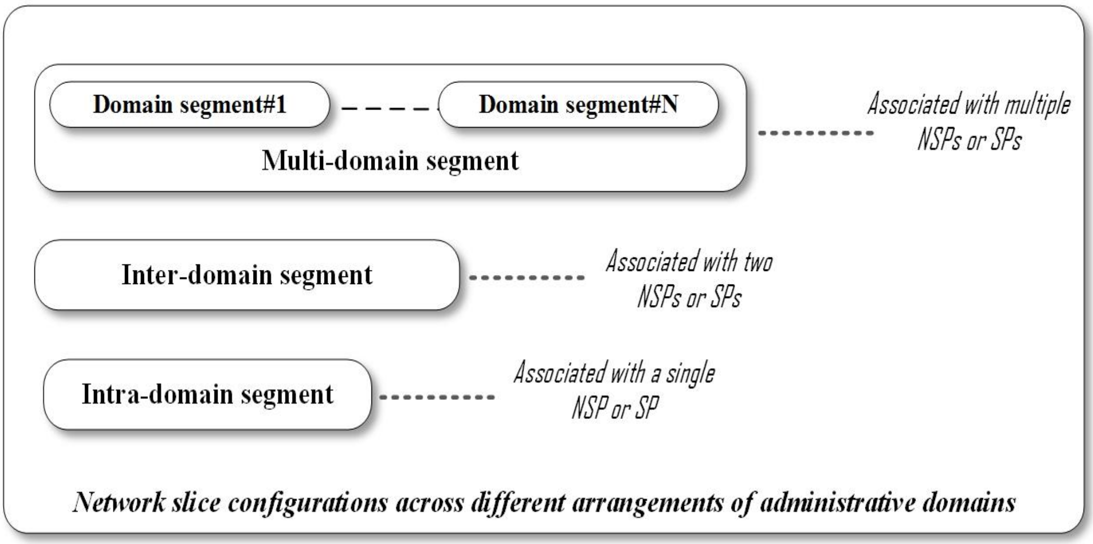
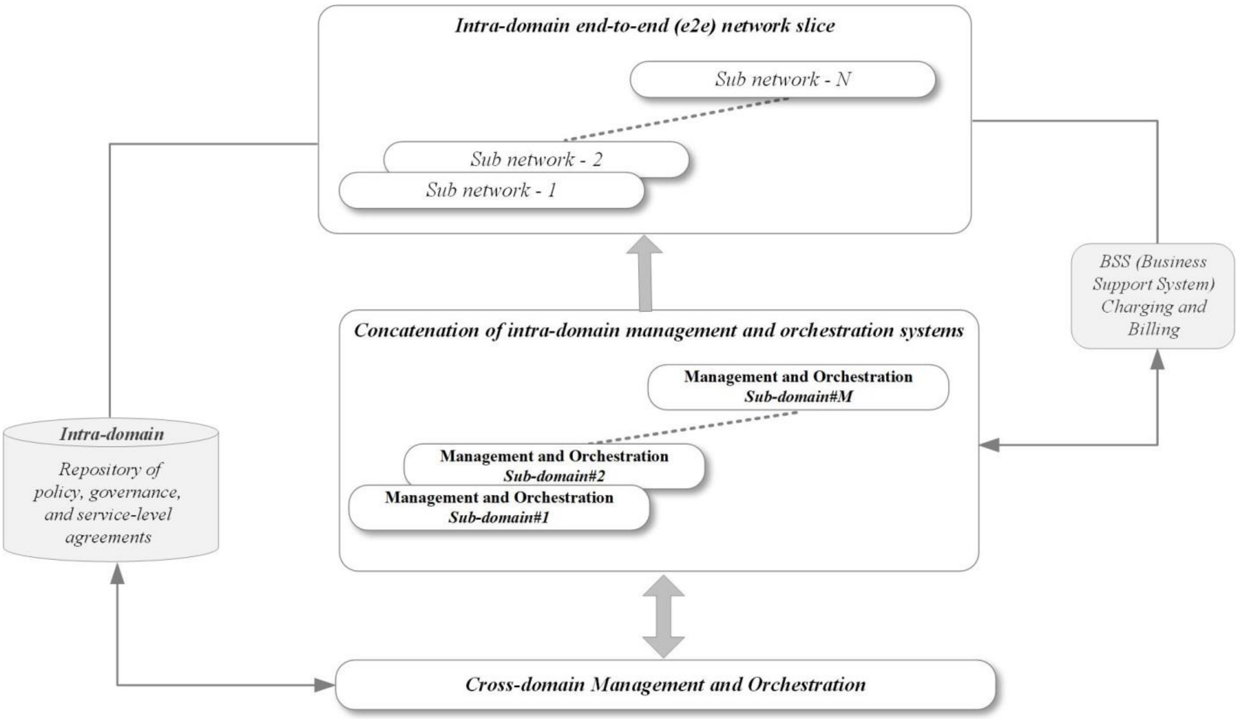

Domain segment#1
Domain segment#N
Multi-domain segment
Associated with multiple NSPs or SPs
Inter-domain segment
Associated with two NSPs or SPs
Intra-domain segment
Associated with a single NSP or SP
Network slice configurations across different arrangements of administrative domains

Fig.12. Categories for network slice configurations

11.2.1 Intra-domain – E2E network slice

Fig.13. depicts sub-domains, if present, within the same domain. Sub-domain management and orchestration systems are depicted as an arrangement of building-blocks from the inter-domain scenario, shown in Fig.14. .



Intra-domain end-to-end (e2e) network slice
Sub network - N
Sub network - 2
Sub network - 1
Concatenation of intra-domain management and orchestration systems
Management and Orchestration Sub-domain#M
Management and Orchestration Sub-domain#2
Management and Orchestration Sub-domain#1
Intra-domain
Repository of policy, governance, and service-level agreements
BSS (Business Support System) Charging and Billing
Cross-domain Management and Orchestration

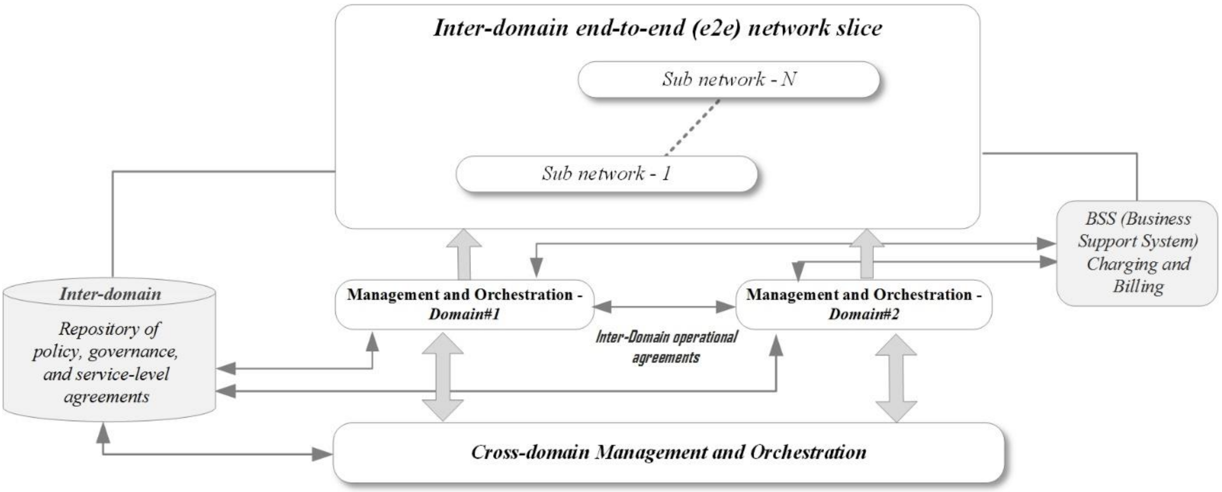
Fig.13. Intra-domain E2E network slice

Fig.13. depicts any number of E2E network slices N that may be instantiated, through coordination across M different sub-domain management and orchestration systems, where N and M are finite integers.

The conceptual building-block, for management and orchestration across more than two sub-domains is derived and applied from the inter-domain scenario depicted in Fig.14.

11.2.2 Inter-domain – E2E network slice

Fig.14. shows the inter-domain scenario.



Inter-domain end-to-end (e2e) network slice
Sub network - N
Sub network - 1
BSS (Business Support System) Charging and Billing
Inter-domain
Repository of policy, governance, and service-level agreements
Management and Orchestration - Domain#1
Management and Orchestration - Domain#2
Inter-Domain operational agreements
Cross-domain Management and Orchestration

Fig.14. Inter-domain E2E network slice

Fig.14. illustrates a conceptual view of the interactions among management and orchestration entities, where two different administrative domains are engaged in the establishment of a network slice.

Fig.14. shows that any number of E2E network slices N, where N is a finite integer may be instantiated, through coordination between the management and orchestration systems in Domain #1 and Domain #2.

11.2.3 Multi-domain – E2E network slice

Fig.15. shows a multi-domain scenario as an arrangement of building-blocks from the inter-domain scenario, shown in Fig.14.

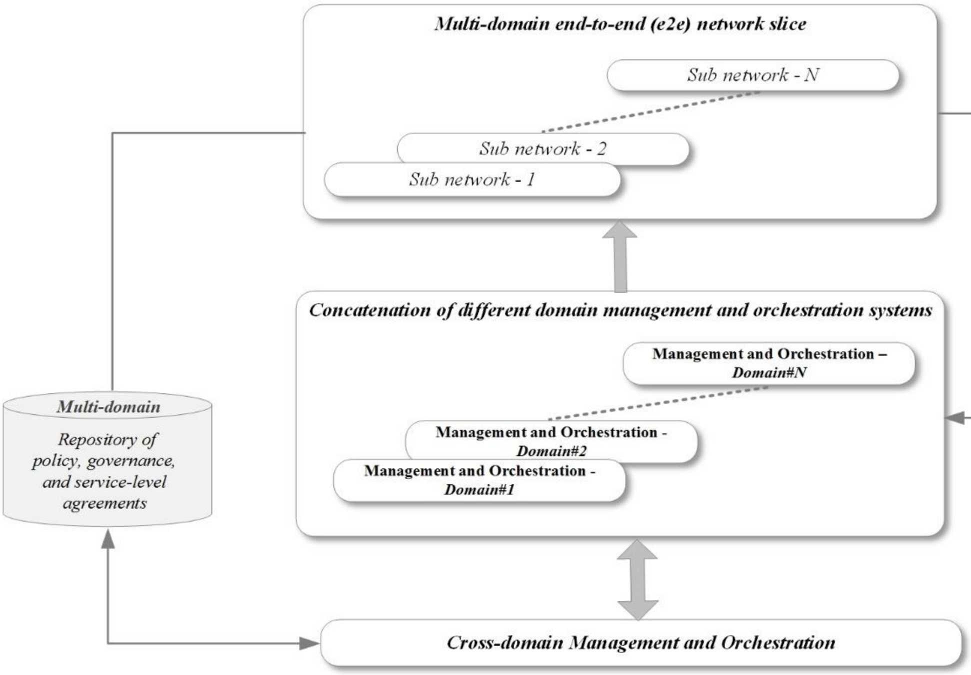

Multi-domain end-to-end (e2e) network slice
Sub network - N
Sub network - 2
Sub network - 1
BSS (Business Support System) Charging and Billing
Concatenation of different domain management and orchestration systems
Management and Orchestration – Domain#N
Management and Orchestration - Domain#2
Management and Orchestration - Domain#1
Multi-domain
Repository of policy, governance, and service-level agreements
Cross-domain Management and Orchestration

Fig.15. Multi-domain E2E network slice

Fig.15. shows that any number of E2E network slices N, where N is a finite integer may be instantiated through coordination across M different inter domain management and orchestration systems, where M is a finite integer. The conceptual building-block, for management and orchestration across more than two domains is derived and applied from the inter-domain scenario depicted in Fig.14. .

The extra complexity of multi-domain (multi-lateral) network slice arrangements over inter-domain (bi-lateral) network slice arrangements introduces additional technical, operational and business challenges. Technical challenges include increased standardisation of the interfaces and models used, and how interconnection can be solved at the different layers. Operational challenges include process, information-model and data-model alignment. Business challenges include a need to align the minimum subset of capabilities that each administrative domain must have, having “standard” hosting/tenancy agreements (similarly to the roaming agreements of today), common settlement methodologies etc.

11.2.4 Federated Network Slicing Service Model

A service model that involves the participation of multiple domains, associated with two or more providers (NSPs or SPs), where a service is rendered for an end-user is depicted in Fig.16. . Each of the providers participates to allocate the appropriate functions and resources to compose a network slice to support “Service X”, which is any given service for an end-user.

The use of a Federated Network Slicing (FNS) approach enables the creation of a network slice, where the network slice orchestrators in each of the participating domains cooperate to create a network slice for the creation of any service for an end-user. The FNS approach avoids the challenges associated with scalability, coordination, complexity, and the preservation of a consistent user experience, in the presence of end-user mobility, across disparate providers.

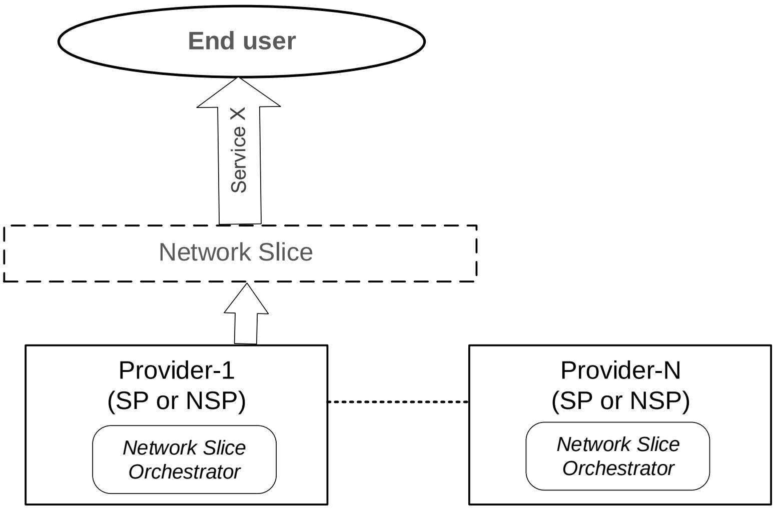


Fig.16. Service model - Federated network slicing

FNS leverages the benefits of forward-looking capabilities, such as NFV/Virtualization.

As in scenarios such as home control/home routing or local breakout, in the FNS approach one provider (SP or NSP), is the designated service provider (P-Hosted) for an end-user. To complete the allocation of all the required functions and resources associated with a given service composition the engagement of other providers (SP or NSP) acting as partners (P-Hosting) may be needed, as required to suit the demands of a given service. For example, these other partners may provide capabilities such as connectivity or infrastructure as a service, using their NFV infrastructure etc.

In addition, a partner provider (P-Hosting) can provide connectivity services allowing the service provider (P-Hosted) to enable services via the partner (P-Hosting) for local access. The partner provider can in this case create a network slice used to provide services for the P-Hosted service provider. Hence a network slice within the partner provider network becomes a part of the service providers network slice (i.e. a network slice subnet).
To set this up the service provider orchestrator will, as part of setting up the network slice and its network slice subnet, send requests to the partner provider orchestrator, which will trigger the creation of the local network slice.

11.3 Inter-domain /Multi-domain / Intra-domain Slice Policy & Governance Repository

This repository stores the following items: Operator’s Slice Policies, Governance, Objectives, Goals, negotiated SLAs/OLAs including compensations, commercial arrangements, responsibility demarcation among (Domains) and related roles and information to be exchanged at touch points.

Those items should be defined, negotiated, prescribed and updated in a flexible manner. Once appropriately combined and translated into desired rules, they will be propagated along the E2E Slice Orchestration path for execution and enforcement purposes.

11.4 The BSS (Charging & Billing) and its articulation with the operator's Slice Policy & Governance Repository

The execution phase includes usage tracking of services and its articulation with the BSS component for reporting, rating, charging per E2E Network Slice and Settlement per Subnetwork Slice provided (when applicable) and enforcement of the compensation rules in case of any SLA clause violation as prescribed in the SLA contract.

12 ENDPOINT/USER EQUIPMENT

12.1 Types

The types of endpoint/user equipment are characterised by a variety of attributes, within three broad categories of interfaces, namely, a) Human-Human (H-H), b) Human-Machine (H-M), and c) Machine-Machine (M-M). A few examples of user equipment or devices that belong to these categories are, smartphones (H-H), robots (H-M) or (M-M), drones (H-M) or (M-M), wearables (H-M), smart objects (M-M) etc. The attributes and capabilities, associated with them are diverse, such as, high-power, low-power, long battery life, low-cost, high performance, latency sensitive, high-reliability, precision sensitive. They are distinguished in terms of diverse media (synchronous, asynchronous, and isochronous) types, such as audio, visual, haptic, vestibular, data streams etc. They may be tethered to a network, either via a wired connection (e.g. Ethernet etc.), a wireless short-range unlicensed connection (e.g. Wi-Fi, Bluetooth etc.), or wireless licensed connection (Cellular for instance).

12.2 Composite access

The availability of different types of RATs (Radio Access Technologies) for endpoint/user equipment to access a network or other endpoint/user equipment characterises composite access. For authenticated customer-side endpoint/user equipment session continuity between different types of access technologies, such as short-range access technologies (e.g. Wi-Fi, Bluetooth etc.) and wide-area access technologies (e.g. different cellular schemes) is desirable. Composite access includes wired access, and different rates of mobility for wireless access. In such scenarios the availability of a diversity of access technologies, allows for optimisation of the utilisation of each access resource, as well as the selection of the most suitable access type for a given session.

Endpoint/user equipment may be connected to several access technologies (including new 5G RATs and LTE) at a given instant, potentially via carrier aggregation, or dual connectivity. The combination of access technologies may involve 3GPP access technologies and non-3GPP access technologies (licence exempt spectrum).

12.3 Heterogeneous access

Within a given type of wireless access technology, the availability of different cell types with different characteristics, which inherently interoperate characterises heterogeneous access. For example, radio network access elements, such as base stations with progressively larger to smaller coverage footprints are referred to as macro, pico and femto base stations respectively. A combination of these types of base stations offers the potential to optimise both coverage and capacity, by appropriately distributing pico and femto base stations within a larger macro base station coverage area.

Since the radio access technology is common across these different types of base stations common methods for configuration and operation such as 3GPP Licensed-Assisted Access (LAA) and optimisation of transmission power levels can be used to manage the allocation and utilisation efficiency of radio resources.

12.4 Cloud radio access

The cloud radio access model can be applied to composite and heterogeneous types of access. Resource offloading from endpoint/user equipment to the edge of the cloud radio access network enables diverse services over a variety of endpoint/user equipment types (e.g. H-H, H-M, and M-M). Other benefits may include energy conservation in the endpoint/user equipment and the ability to offer resource intensive services to the user even if computing or storage resources in the endpoint/user equipment are limited.

13 SECURITY

13.1 Network Layer Security

The 5G system shall support an access-agnostic subscription authentication framework capable of fulfilling the security requirements for both 3GPP and non-3GPP access.

The 5G system shall support subscription identities, for humans and machines/things, according to the international mobile subscription identity plan E.212 and possible future evolutions of this plan.

In 5G, the communication channels on control and user planes shall be integrity and confidentiality protected according to the security requirements building upon 4G and considering the specific role of the multi-access edge in MEC, which may require a combination of end-to-end and hop-by-hop techniques. The security requirements related to MEC [8] shall be respected.

With respect to low latency, the recommendations specified in [8] shall be met.

Trade-offs between ultra-low latency and security, especially with respect to confidentiality protection, may require usage of dedicated very fast cryptographic algorithms. In particular, light-weight, yet secure and trustworthy, cryptographic algorithms (e.g., ISO/IEC [27], [28], NIST [25]), may provide candidates, especially for constrained environments.

E2E network slicing leverages the attributes of network slicing, which is a central virtualisation technology in 5G, which should help flexibly address the variety of use cases with different requirements as well as multi-vendor and multi-tenant network models over a shared hardware infrastructure. The security of network slicing shall meet the general requirements specified in [5]. The most important security-related property of network slicing is the isolation/separation of network slices, including the case where some network functions can be shared.

Multi-layer isolation mechanisms could be introduced in order to reduce the attack surface and scope of impact. Examples include NFVI boundary isolation, isolation of the MANO system, service instance isolation, security domain isolation, VNF isolation, network slice isolation etc.

Fundamental challenges regarding the network slicing isolation are pointed out in [24]. There is a variety of isolation properties which can be satisfied at different isolation levels by using various technologies adapted to the desired isolation levels. The technologies include various software, hardware, and cryptographic mechanisms. The mechanisms may include orchestrator-managed containers, hypervisor-managed virtual machines, and virtual private networks. For each technology selected, it is important to specify concrete security requirements and the corresponding assurance levels.

If endpoint/user equipment can access one or multiple network slices, then, before accessing any network slice, it shall be authenticated to access the 5G system – via primary authentication described in TS 33.501 [31] – and it should be authorised and/or authenticated for each network slice. A common authentication framework (e.g., EAP) should be used for this optional additional slice-specific authentication.

Massive IoT is likely to facilitate new variants of DoS and DDoS attacks on the network infrastructure and connected endpoint/user equipment. The corresponding security threats and recommendations given in [14] shall be respected.

To facilitate early detection and prevention of DoS/DDoS attacks initiated by malicious or infected massively distributed IoT devices, the IoT gateways, managing the devices and connecting them to the Internet, can be used for detecting anomalous or suspicious events. They should cooperate with classical network tools for DoS/DDoS detection and prevention based on analysing the aggregated traffic. New tools that will collect and analyse the RAN traffic may also be useful. This way the network infrastructure and IoT devices can be better protected in a timely manner.

Sensitive data (e.g., in e-health) generated by and transmitted among IoT/M-M devices should preferably be confidentiality, integrity, and anti-replay protected in E2E manner, possibly at the application layer.

In the 5G system, there are new security risks related to software and hardware implementation aspects in virtualised network infrastructure supporting virtualised network functions, network slicing, Service-Based Architecture (SBA), communication between applications, communication between virtualised network functions, and APIs. Therefore, to address the security risks in 5G E2E architecture, in addition to the traditional network security approach based on protecting communication channels and protocols, a holistic approach, involving also computer security and cybersecurity aspects, is needed.

In the context of SBA, a study of the evolution of IPX, IPX provider services, and related security architecture should be performed.

An informative overview of the security challenges and opportunities in SDN/NFV and 5G networks is proposed in [29].

The 5G system should support privacy and security requirements compliant with international regulations, including general data protection and regulation [15], and with the proposal for a regulation on privacy and electronic communications [16].

The 5G system shall support relevant legal and regulatory obligations, including security aspects of Lawful Interception (LI). These legal and regulatory obligations may depend on country or regions.

The 5G system shall support protection of critical infrastructures, including the network infrastructure and 5G use cases such as Industry 4.0, E-health, Public Safety, Transport, Energy, Automotive, and massive IOT. The 5G system should consider compliance with the Directive NIS (Network and Information Security) (EU) 2016/1148 of the European Parliament and of the Council.

13.2 Business Enablement Layer Security

The Business Enablement Layer shall not compromise the 5G security architecture. Some relevant aspects of this layer, e.g., related to roles, responsibilities, and liabilities, are still to be defined.

The Business Enablement Layer may introduce new security requirements. This will especially apply to virtualised implementations (virtual appliances, hypervisors, OS, orchestrators, containers, etc.). See [8].

Since Business Enablement Layer functions shall be realised as virtualised network functions (VNFs) according to the principles specified by ETSI ISG NFV, the security considerations of these VNFs can be in alignment with the guidance published by ETSI ISG NFV-SEC in various aspects [14][18][19][20][30].

Whatever action/decision taken at this Business Enablement Layer, the NSP owning the network infrastructure should be in position to keep its sovereignty on its all network infrastructure. No third party could impose/force the NSP to import third-party software. Furthermore, the NSP, responsible for the network infrastructure, should be always able to remove / delete / isolate whatever VNF already put in production.

13.3 Business Application Layer Security

The Business Application Layer itself is out of scope according to [1].

The Business Application Layer shall not compromise the 5G security architecture.

Cross-application communication should be thoroughly and verifiably tested and strictly restricted to make sure the interface would not open-up new attack surface.

An NSP application shall be integrity protected throughout its lifetime, both in storage and desirably also in running.

Non-NSP applications, including enterprise service applications, vertical service applications, authorised OTT and third-party service applications, may introduce new security risks to the network [8]. The problem should be addressed by using a highly secure virtualization platform with real-time security monitoring of Non-NSP applications (with anomaly detection and attack mitigation) or by running only trusted Non-NSP applications, previously tested according to the adopted quality assurance framework. In practice, security monitoring of Non-NSP applications can be simplified according to the risk analysis and vulnerability assessment conducted by the NSP.

In any case, the NSP should be in position to accept, deny, and remove any proposed application at any time and in any localization (NSP sovereignty on its own infrastructure). No third party or roaming agreement could force an NSP to implement a third-party application, even if the application has been validated by external framework/policy.

13.4 Management and Orchestration Security

Additional work and developments are needed to address the security concerns of some operators regarding the control of their network infrastructures depending on the technology / technical solutions used for virtualisation (e.g., based on containers, orchestrators, hypervisors, micro-services, DevOps). Specifically, opening interfaces/APIs at the orchestrator level, which are in control of the network infrastructure, creates new risk on the control of the operator network. Relevant threat analysis and security requirements are given in ETSI NFV-SEC 014 [21]. In any case, it is the responsibility of operators to make decisions on the virtualisation technologies to be used in their networks.

In general, due to its critical role in network virtualisation and network slicing, virtualisation management software should be highly secure and trustworthy. More specifically, this requirement relates to hypervisors for virtual machines and orchestrators for containers. Virtualisation management software should be free of exploitable software vulnerabilities. Security and trustworthiness of virtualisation management software shall be supported by appropriate assurance levels.

The adopted solutions may depend on the obligations and requirements for critical infrastructures that emerge in certain countries or regions.

13.5 Endpoint/User equipment Security

Subscription data for the authentication of subscriber with the public 5G network shall be stored and processed in a tamper-resistant secure hardware component on the endpoint/user equipment. Solutions under development in 3GPP SA3 are consistent with this requirement. This requirement may possibly be extended under specified conditions to any persistently stored credential data needed for the authentication of subscriber with NSP/SP.

Secret data or privacy-sensitive data for local authentication of a user to endpoint/user equipment (e.g., passwords or biometric templates) should be securely stored and processed on the endpoint/user equipment. For example, this security requirement can be implemented by using the solutions developed by the FIDO alliance [23].

Depending on the use cases, massive IoT may possibly require light-weight (yet secure and trustworthy) cryptographic algorithms that are adapted to power and bandwidth constraints, possibly the algorithms to be recommended by NIST [25] or proposed in ISO/IEC [27], [28].

To proactively protect the network infrastructure from new variants of DoS/DDoS and other attacks facilitated by massive IoT and to protect the IoT devices and the data generated by such devices, it is advantageous for IoT/M-M devices to run on simple, secure, and trustworthy CPUs and operating systems, with verifiable security levels. Preferably, these security levels should also reflect the sensitivity of data generated by such devices.

To proactively protect the network infrastructure from new variants of DoS/DDoS and other attacks facilitated by massive IoT and to protect the IoT devices and the data generated by such devices, authentication mechanisms for remote management of massively distributed IoT devices shall be secure and trustworthy.

A possible future threat is battery exhaustion attacks against devices – either indiscriminately at large scale or targeting individual devices whose disabling has value to the attacker. This is relevant primarily for IoT devices that run on batteries and cannot be easily or frequently recharged. If large scale battery exhaustion attacks become a significant problem, then network-based detection and prevention measures, including traffic analysis and management, may be needed. For individual devices whose availability is particularly important, defence against battery exhaustion attacks may be best implemented in the devices themselves.

The security of cryptographic functions also depends on their implementation, in hardware or software. To resist the so-called side-channel attacks (e.g., timing and power analysis attacks), it is necessary to apply appropriate countermeasures. For example, the requirements given in FIPS PUB 140-2 [22] could be used, with appropriate security level.

If endpoint/user equipment can access multiple network slices simultaneously, then isolation of the slices – and the ability to provide dynamic assurance of this isolation – should be provided by trustworthy software and/or hardware mechanisms based on an immutable root of trust, with appropriate security assurances, to prevent the leakage of (sensitive) data among the slices as well as the disruption of services in the slices. Note that this requirement refers to the mechanisms used to isolate execution and data storage at the endpoint/user equipment.

14 POLICY AND QUALITY OF SERVICE

14.1 General

The 5G system will support a common policy framework along with network policies that allow the endpoint/user equipment to choose the most suitable access network and an access agnostic quality of service mechanism. The 5G system shall support a common quality of service framework.

The common policy framework shall be access aware to enable conformance to service related QoS demands. In scenarios where more than one type of access (e.g. wireless, wired) is available, the choice of access hinges on the optimum (e.g. link conditions, efficiency, performance, policy etc.) suitability to satisfy QoS demands. The non-3GPP access solution could be a subset of the 3GPP access solution.

15 IDENTITY MANAGEMENT

15.1 General

The 5G system shall adopt a systematic approach to manage the following types of long-term identifiers:

- Endpoint/user equipment side:
    - Unique subscription identifier: e.g. IMSI/SUPI, etc.
    - Equipment identifier: e.g., IMEI

    - User (human or machine) identifier: e.g., username or pseudonym

  Unique subscription identifiers are owned and managed by operators. User identifiers are owned by users.
- NSP/SP side:
    - Network element identifier: e.g., a unique ID of an MME/AMF, etc.
    - Network device identifier: e.g., a femto ID associated with a digital certificate, etc.
    - Application identifier: e.g. a code-signing certificate, etc.
    - Service identifier: e.g. MSISDN, a NSP/SP certificate, etc.
    - Human identifier: including administrators, developers, etc.

A role of operators is to bind (e.g., during the registration) the unique subscription identifier to user-related, physically verifiable identifiers such as a username, where the corresponding secret key enables the remote authentication of the identifiers. Therefore, operators are in an advantageous position to perform the identity services also in more general usage scenarios providing new business opportunities.

Accordingly, the operators can provide user-centric identity services. More precisely, the operators can securely provision, manage, and share new identities, credentials, and authorization tokens linked to the subscription identity, provide remote authentication services based on them and control the privacy policies associated with them. The identities of humans may possibly include biometric identifiers. The related opportunities and requirements from GSMA [26] are under consideration in 3GPP. For IoT, to identify devices or things, it may be of interest to use Physical Unclonable Functions (PUFs) as biometric equivalents for integrated circuits. Operators and service providers must comply with privacy regulations on biometric and PUF data including protection against theft and reuse.

The 5G system shall support mutual authentication of an endpoint/user equipment-side identity and an NSP/SP-side identity or of two NSP/SP-side identities.

The 5G system shall adopt a systematic approach to manage short-term or temporary identifiers, for example, 5G equivalents of TMSIs, GUTIs, etc. For un-traceability, except for authorised entities, it should be hard to derive the long-term identifier from temporary identifier(s). For un-linkability, except for authorised entities, it should be hard to track the movement of the same endpoint/user equipment based on temporary identifier(s).

The use of strong user authentication, which is not based only on passwords or PINs, is encouraged. For example, the solutions promoted by the FIDO alliance [23] may be useful for this purpose.

Long-term or permanent identifiers (e.g., IMSI/SUPI, IMEI, username) shall only be visible to authorised entities that need them for providing their function. They shall be stored securely and should not be transmitted in the clear. If they are transmitted encrypted, then the encryption shall be randomized for privacy reasons, to avoid linkability. The encryption/decryption key should be stored securely, and the encryption/decryption operation should be executed in a secure environment.

16 LIST OF ABBREVIATIONS

3GPP	Third Generation Partnership Project
4G	Fourth Generation 3GPP system
AAA	Authentication, Authorisation and Accounting
AD	Autonomic Domain
AI	Artificial Intelligence
AuN	Autonomic Networking
AM	Autonomic Manager element, which is also referred to as a DE

API	Application Programming Interface
BSS	Business Support System
cDE	Centralized DE
CP	Control Plane
CPE	Customer Premises Equipment
CPS	Control Plane Service
CU	Centralised Unit
D2D	Device-To-Device
dDE	Distributed DE
DDoS	Distributed Denial of Service
DE	Decision-making Element, which is also referred to as an AM element
DPL	Deep Learning
DoS	Denial of service
DT	Domain Type
DID	Domain Identifier
DU	Distributed Unit
EAP	Extensible Authentication Protocol
E2E	End-to-End
EM	Element Manager
eMBB	Enhanced Mobile Broadband
ETSI	European Telecommunications Standards Institute
FCAPS	Fault, Configuration, Alarm, Performance and Security Management.
FMC	Fixed Mobile Convergence
GANA	Generic Autonomic Networking Architecture
GUTI	Globally Unique Temporary Identifier
GSM	Global System for Mobile Communications
GSMA	GSM Association
H-H	Human to Human
H-M	Human to Machine
IaaS	Infrastructure as a Service
IEEE	Institute of Electrical and Electronics Engineers
IMEI	International Mobile Equipment Identity
IMS	IP (Internet Protocol) Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IPX	Internetwork Packet Exchange
KP	Knowledge Plane
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LTE	Long Term Evolution
M-M	Machine-to-Machine
MANO	Management and Orchestration
ME	Managed Entity
MEC	Multi-access Edge Computing
mIoT	massive Internet of Things, typically referring to 5G IoT
ML	Machine Learning
MMTEL	Multimedia Telephony
N6	3GPP interface between the 5G core network and a Packet Data Network
NaaS	Network as a Service
NE	Network Element
NFV	Network Function Virtualisation
NFVO	NFV Orchestrator
NFVI	NFV Infrastructure

NSP	Network Service Provider
OLA	Operations Level Agreement
OSS	Operation System Support
OTT	Over-The-Top
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PDV	Packet Delay Variation
PELR	Packet Error Loss Rate
PNF	Physical Network Function
PUF	Physical Unclonable Function
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
SBA	Service-Based Architecture
SDO	Standards Developing Organization
SDN	Software Defined Networking
SLA	Service Level Agreement
SON	Self-Organising Network
SP	Service Provider
SUPI	Subscriber Permanent Identifier
TMSI	Temporary Mobile Subscriber Identity
UP	User Plane
UPS	User Plane Service
URLLC	Ultra-Reliable Low Latency Communication
VIM	Virtualized Infrastructure Manager
VNF	Virtualised Network Function
VNFM	VNF Manager
X-Haul	Flexible, heterogeneous access fronthaul and backhaul

17 ANNEX: AUTONOMIC NETWORKING (AUN)

17.1 Design principles for AuFs (Autonomic Functions)

The following is a list of related requirements, with respect to NEs (Network Elements) and NFs (Network Functions) in the realm of AMCs, based on [42]*:*

- ❖ The various specific autonomic decision-making functions referred to as DEs (Decision Elements) should be defined and characterized for closed-loop operation for their individually assigned MEs (Managed Entities). For example, a DE is an autonomic function that enables the implementation or realization of a control-loop for its assigned ME.
- ❖ The components of DEs should be viewed as software logic or algorithms that may be implemented as standalone processes (e.g. microservices) or combined in some fashion at runtime as a single process.
- ❖ The logic in a DE should be modular, with the autonomic manager components being runtime entities that are re-loadable and replaceable software modules. These software modules are such that they can be re-loaded into NEs and NFs to enable fast control loops within NEs and NFs, while enabling slower control loops outside of NEs and NFs, within the management and control system. This NE/NF behaviour complements those of DEs in NEs and NFs through policy control.

  Note: The ability to re-load and replace DEs with best-in-class autonomic decision-making elements that exhibit better algorithmic behaviours over time is beneficial for SPs, since DEs serve as instruments of innovation in terms of artificial intelligence and cognitive capabilities for autonomics for differentiation across different DE implementations.

- The holistic nature of a multi-layer hybrid AMC model shall be characterized in terms of key levels of abstraction at which autonomic control loops and the associated DEs can be implemented to interwork together in the form of multi-layer autonomics, through a framework of AMC oriented networks and services.

- The multi-layer hybrid AMC model shall be generic, interoperable, and specified in terms of key functional blocks and reference points that enable an implementation and instantiation of generic autonomics in target network architecture and its associated management and control architectures.

- The AuFs should be clearly specified such that the value of each AuF is described in terms of the presence of these AuFs within NEs and NFs or within the realm of management and control systems that provides a measure of the management and control intelligence capabilities of entities that embed AuFs.

  For example, the types of AuFs include:

QoS-management-DE	*Security-management-DE*	*Mobility-management-DE*
Fault-management-DE	*Resilience & Survivability-DE*	*Service & Application management-DE*
Forwarding-management-DE	*Routing-management-DE*	*Monitoring-management-DE*
Generalized Control Plane management-DE		

Fig.17. Types of AuFs

The NEs, NFs, and autonomic components in the network and the associated management and control system should be enabled with auto discovery of resources and information in the network, through a scalable federated arrangement of computing resources.

For example, auto discovery capabilities to discover resourced and information in the network can be accomplished through “publish/subscribe/query and find” schemes [42].

- The interaction and coordination among the AuFs, such as DEs, should consider both the hierarchical nature of control loops and the peer-to-peer DE-to-DE horizontal interactions that may be required in certain autonomic use cases requiring DE-to-DE algorithms, implemented in a distributed fashion [42].

- A translation function, such as an MBTS (Model Based Translation Function), may be required as an intermediary layer between AuFs, in the realm of management and control system, NEs and NFs, whether these are physical or virtual.

  Such a translation function is useful for translating technology specific or vendor specific raw data into a common data model for use by higher level AuFs, based on a well-defined and shared information/data model. The translation function may also include a translation of monitored data from NEs and NFs into ‘knowledge’ that can be used by higher level AuFs in their decision-making process.

17.2 E2E decision making process and governance

The KPs complement lower-level autonomics in NEs and NFs, for a holistic multi-domain state correlation and adaptive resource re-programming, for specific network segments and domains, such as access, backhaul, and core networks etc. [42] [48] [49]:

- E2E autonomic service assurance for E2E services, including network slices, should be achievable through a federated autonomic management and control framework across technology and administrative domains engaged in 5G service delivery, where such 5G domains exhibit autonomic and cognitive capabilities to minimize or avoid human intervention for management operations.

- All the DEs in the AMC framework, within an E2E architectural framework, should be governable for a configuration in either an open-loop or closed-loop control mode, for adapting to policies or other inputs provided by automated tools available at human interfaces for a generation of data associated with the network or specific services.

- The NSP should be able to test, certify, trust, and validate any DE.

- The E2E architectural framework should incorporate an AMC framework that is enabled in terms of cognitive algorithms, including artificial intelligence, for DEs governed by policies and operational objectives. These policies and operational objectives for the SP or NSP are available for modification through human intervention to control any design time and runtime for an optimized adaptation to changes that are sensed by the AMC framework.

- Each DE within the E2E architectural framework shall be replaceable, upgradeable, and controllable by the SP or NSP at any time during the lifecycle of the network, especially in the case of the KP DEs [42].

- AuFs should provide certain guarantees for SPs or NSPs in terms of their management and control operations, through attributes such as 'trust and confidence', stable and self-coordinated non-conflicting management and control behaviour, across different network domains (e.g. RAN, X-Haul/Backhaul, Core network etc.).

- A network segment or domain with the E2E architectural framework should have the capability to provide a fine-grained decomposition of both centralized/slow control-loop DE (cDE) and distributed/fast control-loop DE (dDE).

  These types of DEs have a specialized scope of 'management and reasoning' that distinguishes the type of DEs. For example, the different types of DEs include those shown in Fig.5. , which does not preclude other types. These DEs could be deployed as a specific service in a cloud-native environment, for a specific management and control scope or domain [42], which defines various types of DEs and their scope of "management and reasoning" in their operations.

- The AMC framework, within the E2E architectural framework should define an 'ownership' model, during the design phase to identify which DE autonomic function is allowed to interact with certain assigned or owned MEs, as well as how the DE can interact with other MEs associated with another DE, in order to guarantee a conflict-free coordination, during the instantiation or runtime operation of a DE.

- The E2E architectural framework should incorporate a logically centralized and shared data or knowledge repository with federated computing resources that support common semantics in the form of a data model, which is accessible for data analytics driven entities, such as DEs, through open and well-defined APIs.

  The availability of support a 'Publish-subscribe' model for enabling some entities to publish information in the shared data or knowledge repository or to subscribe to receive data from this repository should be considered.

  The availability of support for a 'Query and Find' model for enabling some entities to discover information of interest to these entities in this repository should be considered.

Note: To enable auto-discovery of resources and information in the network, autonomic networking capabilities promote an implementation of ONIX (Overlay Network for Information Exchange) [42], where a system of federated information computing resources serve as a real-time inventory that supports a 'Publish/Subscribe/Query-Find' model.

- The AMC framework, within the E2E architectural framework should specify and make available a translation scheme or function to enable communications between a logically centralized and shared data or knowledge repository, associated with each domain, among a federation of domains, at cross-domain reference points, where each domain may be interpreted a technology oriented or administration oriented.

- The AMC framework, within the E2E architectural framework shall specify a hierarchical and horizontal model (sibling and peering relationship), between DEs located either within the same domain instance (single domain scenario), or across domains, such as in the case of two or more federated domain instances.

  This model enables any given DE that is hierarchically arranged, with respect to other DEs to exercise policies for controlling other DEs that are siblings, for managing their behaviours, where 'situation management' is subjected to a delegation, synchronization, and escalation process in a hierarchical arrangement of DEs.

On the other hand, this model supports a federated arrangement of DEs, where the DEs are positioned at certain points in the infrastructure elements consisting of NEs and NFs, or in domain specific KPs. In this case, a coordination of DEs should be required in order to guarantee a conflict-free autonomic management and control operation across network segments and domains, such as in the case of access network, backhaul, and core network etc.).

17.3 Common and generic information at the federated AMC reference point

- The type of domain, to which the DE is associated, whether it 'technical' or administrative', and whether the DE is enabled to share common and generic information, subject to trust and security policies should be specified.

- The discovery of a 'technical domain' to which the DE is associated for its capability to orchestrate and autonomically manage and control should be specified based on the GANA model in [42]. The GANA model includes the concept of a 'capability model', self-description, and publishing by a DE through which information about a 'technical domain' is encapsulated.

- A DE should aggregate the capabilities of its assigned MEs, and advertise the associated descriptive information as required, subject to trust and security policies. For example, a Function-Level-Routing-Management-DE in a routing device running two protocols, such as OSPF or BGP, would indicate these routing protocols for the associated ME (routing device) under the control of the DE, in the 'capability model' description.

- Administrative domains may have domain identifiers assigned by an associated governing authority following a scheme of choice.

- Other types of common and generic information should include the following types of information:

  - *Synchronization of actions across multiple KPs [42]*
    This is required, for example, to realize an effective E2E federation of Orchestrated Closed-Loop Security Management and Control (adaptive security enforcement and defense) in network

infrastructure segments and across multiple domains (Technologically and/or administratively diverse domains).

Examples of such domains are network segments (domains) such as Radio Access Network (RAN), X-Haul Transport Network (i.e. Fronthaul, Midhaul, Backhaul, etc.), “Multi-Access Edge Computing” (MEC) site or Core Network.

E2E autonomic security management and Control should be achievable through a federation of KPs for the various network segments (domains) associated with a given E2E scope. In this case, each KP policy controls the AuFs running in certain NEs/NFs within the network segment governed within the scope of the associated KP. The AuFs that implement the security policies for self-protection and self-defense of associated NEs/NFs or a secure network zone are required to realize ‘fast control loops’, within the NEs/NFs, in accordance with the generic autonomic networking principles [41].

- *Security event information (regarding a description of a detected security incident)*
  An example is detected threats that may impact a peer domain, which could trigger an investigation of the detected threat that is identified by the collaborating KPs, and may result in the KPs collaboratively negotiating an adaptation strategy (self-adaptation without human involvement) for adjusting security enforcement policies that each KP then applies to realize self-protection and self-defense for its associated network segment/ domain against the detected threat(s).

  For example, there may be some security threats detected in the access network domain by the KP for the access network that could have impact on X-Haul transport network domain as a peer domain or may have impact on the core network as the peer domain in terms of impact scope.

- *Trust model (e.g. a reputation-based trust model) between the autonomic management and control (AMC) administrative domains*
  An example of such a trust model would be across autonomously managed and controlled domains, with the associated network infrastructure segments and their associated KPs, where each particular network segment has a KP.

- *Security related SLA violation detection*
  The detection of an SLA violation, requires the associated KPs to initiate a resolution through a collaboration across the KPs to resolve the SLA discrepancies by reacting to resolve the detected discrepancies for an alignment with the configured SLA clauses in the SLA contracts that were established by the associated domain owners or stakeholders/partners

17.3.1 KPIs exchanged at the federated AMC reference point

The federated AMC reference point should include high-level KPIs that require to be exchanged, when two or more DEs are associated with this reference point referred to as the DE-to-DE interface across multiple network elements at certain points in the network topology [42]. In addition to [42],] [47] [48] [49] provide illustrations for a federation of KPs for various network segments as autonomic management and control domains.

Other types of KPIs may be included.

The KPIs shall include the following types of information:

- Trust levels as a measure of trust worthiness.
- Threat counts of potential impacts to a peer domain, and the severity of the threats.

- Aggregate states of the domain, such as access network, backhaul, and core network etc., in terms of workload or load levels.

- Weights that are a measure of the willingness of a peer domain to deliver certain services, such as transport services.

- Several other types of KPIs that function as security indicators that can be certified and specified.

17.3.2 Federated AMC reference point

The federated AMC reference point should support the following types of information exchange:

- The DEs exchange information such as DT (Domain Type) and DID (Domain Identifier) for a verification of security and trust policies, and for behaviours for the manner in which they configure their associated MEs, for a fulfillment of the required network behaviours across domain boundaries.

  The exchange of information such as DT and DlD may be restricted to the Node-Main-DEs [42] of NEs and NFs, which discover and exchange such information on behalf of lower level DEs, which then use this information for corresponding behaviours that determine the way these lower level DEs provision an autonomic management and control service across domain boundaries [42].

- The MBTS (Model Based Translation Service) [42] instance may be used to translate the information at the KP between logically centralized and shared data or knowledge repository belonging to different domains or between DEs, such as KP DEs belonging to different domains. The MBTS instance may be used to translate information retrieved within a dedicated domain to information representation and presentation format(s) required by a peer domain.

- The federated AMC reference point between DEs within different administrative at the same level of autonomic abstraction may embed security and trust mechanisms embedded in the DEs or those in a separate function or scheme.

- A dedicated reference point between administrative domains may be specified, in cases where the GANA model is utilized for any specific reference architecture, such as an access network, or a core network [48].

- Interworking functions may be specified in some cases, with MBTS and broker functions, which are used to describe distributed peer-to-peer KPs within different administrative domains.