International Network Generations Roadmap (INGR)

An IEEE 5G and Beyond Technology Roadmap

Systems Optimization

1st Edition White Paper

FutureNetworks.ieee.org/roadmap

International Network Generations Roadmap (INGR)

Chapters:

- Applications and Services
- Edge Automation Platform
- Hardware
- Massive MIMO
- Satellite
- Standardization Building Blocks
- Millimeter Wave and Signal Processing
- Security
- Testbed

White Papers:

- Connecting the Unconnected
- Deployment
- Energy Efficiency
- Systems Optimization

Table of Contents

ABSTRACT

Fifth generation (5G) networks set the first step from evolutionary to revolutionary networks. Use cases driving this transition for 5G networks focus on the need to support heterogeneous traffic such as eMBB, URLLC and mMTC. On the software and control side, 5G and beyond networks are expected to support SDN and NFV technologies and will leverage the merging of communication and computing through the "wireless edge". With the deployment of novel applications and the expected increase in their usage and demand, the scope of innovation within future networks will be governed by: (a) limitations and boundaries of available resources; (b) limitations of the adaptability of legacy solutions (scalability and flexibility); (c) limitations of available decision making entities (network slice orchestrators and SDN-controllers will not be enough); and (d) lack of intelligent management and control solutions for multi-variate optimization.

Technologies are available for efficient use and self-adaptive optimization of resources using enablers such as AI-powered autonomic control loops. With ever increasing complexity expected for beyond-5G networks, there is a necessity for novel design, planning and operations paradigms. There is a need for assessment of legacy tools vs new Artificial Intelligence solutions for applicability to systems optimization, and a need for introduction of novel methods to model and study the behavior of highly complex systems developed for the realization of 5G and beyond networks. The goal of this working group (WG) is to assess complexity challenges for the 5G era and beyond, explore novel design, planning and operations techniques for networks and services, and explore intelligence sciences to create the roadmap of the IEEE Future Networks Initiative (FNI) Systems Optimization WG.

Key words: Systems Optimization, Traffic Variance, Control Variance, Service Variance, Confluence, Dependency, Complex Systems, Self-Organizing Networks, Self-X.

SYSTEMS OPTIMIZATION WG WHITE PAPER

1. INTRODUCTION

This white paper gives a broad summary of what one can expect from the more in-depth roadmap effort for the broad topic of systems optimization. It describes a high-level perspective and projection of the topic's technology status, focusing on the challenges and gaps to be explored and reported in the 2020 edition of the IEEE INGR roadmap. The scope and stakeholders are summarized. Any expected linkages among the other INGR roadmap working groups (WGs) are presented.

NOTE: This working group roadmap does not endorse any solution, company, or research effort.

The idea of Self-Organizing Systems (SOSs), although long known from domains such as physics, chemistry, and biology, has gained interest to be applied to technical applications. The reason for this is a paradigm shift from monolithic systems or systems with a small number of components to large networked systems. This paradigm shift is driven by the technological advancement and the emergence of pervasive systems integrating information processing into everyday objects and activities. Such systems can use the view of multiple nodes to come to a massively distributed view of a technical process, where the fusion of several node measurements potentially leads to a more extensive, more accurate, and more robust observation. Note in SOSs, the system as a whole is autonomous, however, individual subsystems may have external control, potentially from a central unit within the system [1].

Realizing autonomous systems requires a *control paradigm* that copes with the complexity of such a solution. A promising approach to attack this problem is the *principle of self-organization*, where the control is distributed throughout. Through the definition of the behavior in local interactions, it is expected that the overall system shows an *emergent behavior* with properties like robustness, adaptability, and scalability as well as complex order. Previous work on self-organizing systems and hierarchical vs. distributed control provides background in this area (see Annex A).

Designing, controlling, and optimizing such SOS is highly challenging due to mutually conflicting goals of various entities, the number of variables, nonlinearity of the problem, local optima, and limited observability and controllability of the plant or environment or process to be controlled. There is no general methodology yet explaining how to design such a system or how to concisely validate it. However, to communicate these problems, potential methods, and obtained results, it is important to have a common understanding of such SOSs.

It is the aim of the IEEE Future Networks Initiative Systems Optimization Working Group to form a scientific community to define SOSs, identify key problems, and provide solutions based on various tools ranging from machine learning and autonomic/autonomous decision making solutions, to complex systems theory, and many others.

1.1. CHARTER

Future systems will be highly distributed fabrics of compute, intelligence, and networking interconnected at multiple levels. "Fabric" here is used in the general sense of a framework that stitches multiple constituents of a collaborative cluster together. This can be applied broadly, such as the basic fabric of a society, an organization, etc. that enables it to function successfully, or narrowly, such as a domain specific fabric (e.g. a compute mesh, or energy grid or a networking fabric) on the other extreme, or something in the middle such as a smart city fabric formed of heterogenous smart clusters stitched together.

The Systems Optimization Working Group is dedicated to identifying key problems of future highly complex and self-organizational networks reflecting future systems, to generate solutions to achieve self-organization, and to demonstrate the proposed features within the scientific community.

The following have been identified as key areas of exploration to support optimization of future highly distributed fabrics:

- Dynamic fabric allocation with (near) real time discovery and peering of heterogenous resources contributed by disparate providers
- Dynamic semantics discovery and negotiation at points of attachment between peer entities
- Distribution and federation of intelligence across disparate contributing entities
- Self-optimizing techniques for autonomous/autonomic system behaviors

1.2. Scope of Working Group Effort

The Systems Optimization working group within the IEEE Future Networks Initiative will address:

- modeling of control of complex networks of self-organizing systems,
- identification of the key problems for control of such networks,
- development of new solutions to achieve network self-organization, application of machine learning (ML)/artificial intelligent (AI) solutions, and/or design and generation of novel ML/AI tools specifically designed for SOS problems,
- demonstration of these features and solutions within the scientific community,
- collaboration with industry and standards community.

Self-Organizing Systems

The basic principles of self-organization have been investigated by several researchers for decades. A SOS, hereby, consists of a set of entities. The management and control in such systems is completely distributed, i.e., each participating subsystem has its own control process.

As depicted below, SOSs are not necessarily a separate field in science but can be found in multiple fields. Hence, they are highly interdisciplinary by nature.

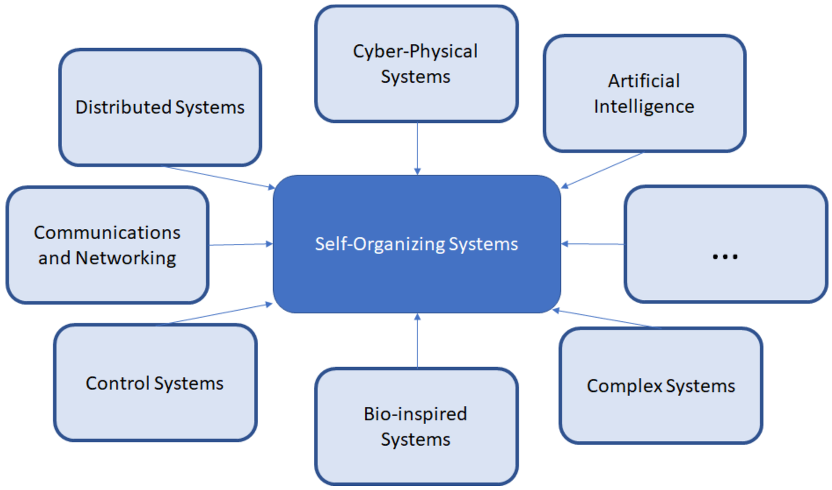

Figure 1: Examples of Self Organizing Systems

SOSs are typically dynamic and structures within them emerge through interactions of the system's entities. Such structures are called dissipative [2].

Complex systems theory looks exactly for systems showing such dissipative, nonlinear behavior, where complex systems are defined as systems of many components which are coupled in a nonlinear fashion. Typically, research on complex systems is based on models using the mathematical techniques of dynamic systems, which include differential equations, difference equations, and maps. Research related to complex systems significantly overlaps with the concept of self-organization. Within the context of complex systems theory, management and control mechanisms for dynamic, highly scalable, and adaptive systems are required and self-organization is accepted as a foundation to achieve solutions for these mechanisms.

<u>Self-Organization and Emergence</u>

Self-organization is typically seen as a process in which the structure and functionality of a system emerge solely from interactions among its entities without any external or centralized control. The system's entities interact locally and exchange their observations without any reference to a global pattern. The interaction of single entities finally defines the behavior of the global system. This is known as emergence of patterns or system behavior.

Emergent behavior of a system or emergence is provided by the evidently significant collaboration of system entities to reach capabilities of the overall system (far) beyond the capabilities of a single entity. In other words, emergence refers to the arising of structures, patterns, and properties during the process of self-organisation in complex systems. Emergent phenomena arise from entity-level interactions and processes and occur on the system level.

Within the context of self-organization and emergence, the following aspects, amongst others, need to be considered:

- absence of external control,
- adaptation to and learning from changing conditions,

- complexity,
- control hierarchies,
- dynamic operation,
- equilibria and local optima,
- self-X capabilities such as self-awareness, self-configuration, self-monitoring, self-optimization, etc.

For example, one of the approaches under study employs the principles of 'Emergence' to drive confluence of Cloud and Communications systems with Emergent Intelligence, wherein Emergent Intelligence is the order that results from the non-linear interactions between components at different levels of a self-organizing Cloud and Communications grid system such that overall intelligence of the system is greater than the sum of the individual intelligent components [3]. At IEEE FNI, we are engaged in assessment of future scenarios that are difficult to address (if not impossible) with traditional approaches. Distributed Security and Mobility Management across heterogeneous RATs for ultra-reliable low latency applications, and optimization of handoffs for the same is one of the most pertinent areas that need immediate attention.

Optimization

In a multi-objective problem, the optimum is not unique, but there can be many optima (Pareto optima) [4]. Systems optimization techniques are closely tied with the desired KPIs (Key Performance Indicators) for a specific system or set of applications. These KPIs and desired applications are dependent on various factors such as underlying infrastructure and policies set forth by regulators. The KPIs (e.g., cost, bandwidth, delay, packet loss, security, system control) are tied with a specific type of application that will be supported for a typical vertical (e.g., first responder, smart city, tactical networks, intelligent transportation system). The underlying architecture will need to optimize various aspects of the systems based on the dependency among various components and sub-systems. For example, a virtual reality type application mandates high bandwidth (~10 Gbps traffic), low latency (~few ms latency) and high system control. In order to support this type of application, one needs to simultaneously orchestrate New Radio (NR) on the RAN and processing at the Edge Cloud to support ultra-low latency.

1.3. Linkages and Stakeholders

Systems optimization techniques can be applied to various parts of next generation network eco-systems as there is dependency among various components. Thus, the Systems Optimization WG will need to interact with various other INGR WGs, e.g., Applications and Services, AI/ML, Connecting the Unconnected, Edge Automation Platform, Security, Standardization Building Blocks, Testbed to understand the dependency among various components in the 5G eco system. Systems Optimization WG will interact with Standardization Building Blocs WG to discuss potential standards that can be pursued, with Security WG on context-aware/adaptive security, with Energy WG on tiered models for optimization of energy costs, and with the Testbed WG to discuss requirements on testbeds for systems optimization.

Finally, Systems Optimization WG will need to interact with Applications and Services WG and Connecting-The-Unconnected WG to discuss the application of systems optimization principles in different use cases. The Systems Optimization working group can focus on a specific vertical (e.g., smart city, first responder, eHealth, V2X, UAV), but aims to provide use case-agnostic solutions. The WG members will then focus on the underlying architecture and find the gaps that can be taken care of by systems optimization. It is anticipated that the WG will look into end-to-end systems and develop holistic solutions.

Members of Systems Optimization WG will need to interact with other standards, namely 3GPP, ETSI, TMF, NGMN, BBF, ITU-T and IETF and accompanying communities such as O-RAN and LF. Some of the systems optimization techniques and algorithm can be applied to various architecture frameworks developed as part of 3GPP and other major SDOs/fora and open source communities. Many of these systems optimization techniques can also be discussed in IRTF and other groups. Previous or current work that may be of particular interest include the Systems of Systems Engineering (for details see Appendix B) and ETSI Generic Autonomic Networking Architecture (GANA, for details see Appendix C).

NGMN and 3GPP have standardized the Self-Organizing Network (SON) concept, part of the broader vision of Self-Optimizing Systems/Networks which incorporates operational principles of network self-optimization, but with scope focused on the RAN network segment only. Efforts to broaden the scope of closed control-loops driven management and control of network resources, parameters and services beyond RAN and into other network segments are underway in major SDOs/Fora such as ETSI, NGMN, 3GPP, ITU-T, BBF, and TM Forum. Emerging standards such as ETSI GANA (ETSI TS 103 195-2, see Appendix C) include principles for abstraction levels for control-loops designs, hierarchical control-loops and nesting, federation of control-loops and decision-making components for autonomics, and use of AI algorithms for the cognitive capabilities in autonomics.

The Systems Optimization WG will explore the use of emergence to address full-stack self-organizing systems, i.e., multi-layer and multi-domain organization and optimization of multiple stacks comprising of heterogeneous radio resources (e.g., 3GPP and non 3GPP RAT), fixed access and transport resources (e.g., optical wavelengths), and compute and store infrastructure resources contributed by disparate service providers. This objective can take advantage of the features provided by IDN (Intelligence Driven Network), which complements SON by deriving current softwarization techniques, such as SDN, to exploit intelligence information and methods.

Stakeholders could include various verticals that are interested to improve efficiency, flexibility, and control latency for their operation during the deployment phase. These verticals could realistically include operators, enterprise networks, first responder, public safety, and tactical network community and app developers. Disaggregation of RAN networks and the associated complexity of control (for details see Appendix D) is a natural area of application.

2. CURRENT STATE

Current technology, e.g., in wireless systems, requires only a fairly simple set of interactions among components of a system, with intelligence and resources often very centralized.

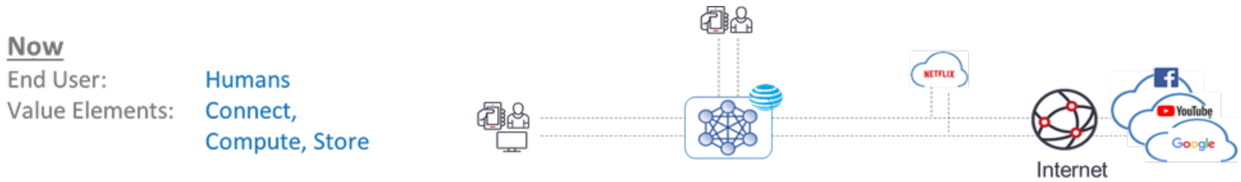

Figure 2: Current State

Previous generations of wireless systems have been an incremental upgrade in bandwidth and performance catering to similar user environments as the previous. 5G, however, is anticipated to be revolutionary in many aspects as extreme radio characteristics/technologies afforded with 5G finally puts wireless on par with wireline broadband ushering in an era of pervasive connectivity. 5G is the first wireless release with the potential to extend Service Provider reach beyond mere broadband connectivity. In today's eco system, end users are humans that use connect, compute, store paradigm to deliver and share. There is less dependency among the system components. Figure 6 shows the current state of interaction among various system components.

Service Variance

The user landscape anticipated with 5G is extremely complex, as it is expected to serve three different extremes, namely enhanced Mobile Broad Band (eMBB), massive Machine-Type Communications (mMTC), and Ultra-Reliable Low-Latency Communications (URLLC) services. An important observation here is that traditional architectures employing straight virtualization of monolithic network functions, for statically preconfigured services, will not be effective due to high degree of variance in service characteristics and dynamicity of scale expected with 5G in the domains of time, space and QoE/QoS. Given the high degree of heterogeneity and sparsity of resources, simultaneous guarantee of these service requirements in an ad-hoc and on-demand manner, and higher resilience requirements to natural and human attacks and failures will lead to enhanced complexity, which needs to be managed effectively and intelligently. Zero-touch service implementation will be expected.

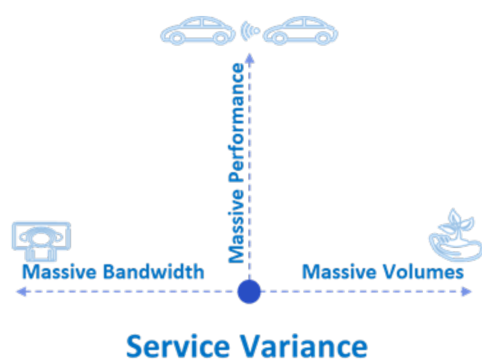

Figure 3: Service Variance

Traffic Variance

Current networks are configured for user-to-content traffic with centralization of data in content centers located in metro and rural areas. Internet access points and cache locations placed closer to the points of content consumption is a perfectly sound model for content delivery with effective use of capacity over metro and long-distance backbones, and low latency between users and content processing locations.

Though, both the nodal interconnection and computational positioning aspects of this model are challenged by the ultra-low latency control expected with 5G and the order of magnitude higher volumes of data exchange expected with autonomous device swarms.

Mobility presents an additional challenge as the service (i.e. the service anchor) must move with the user to maintain consistent service performance, particularly for highly reliable, low-latency control. Future traffic patterns are expected to evolve with increasingly complex connectivity at each step, from today's centralized environment to being distributed toward the Access Edge, and ultimately leading to meshed connectivity to support complex user environments of the future.

Figure 4: Traffic Variance

Control Variance

Optimal control of dynamically responsive compositions of heterogeneous intelligent components (which may in turn be autonomic systems within) is a key challenge for future systems. Autonomy and autonomics in the context of automated management have been researched in academia with selective industry participation for over a decade now. Nevertheless, this area remains with open research challenges. Key among these challenges are placement and federation of control loops for a robust control hierarchy, and data ownership and federation across multiple control jurisdictions. Efforts on addressing these challenges will consider standardization work on autonomics and associated frameworks such as the ETSI GANA work (see Annex C).

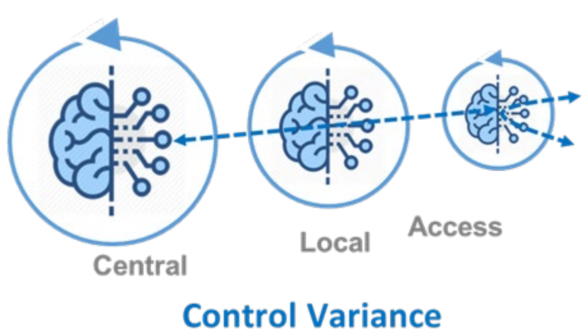

Figure 5: Control Variance

3. Future State

We believe that in future there will be pervasive distribution of computation and storage resources, with the accompanying requirement for pervasive connectivity to allow interaction between distributed, intelligent systems. We also believe that this new wave of pervasive connectivity will lead to pervasive automation as ML techniques applied to robots and smart devices evolve from simple regression to complex decision making – that day is not too far when we'll be immersed in unthinkable experiences enabled with swarms of autonomous devices all around us

- as the end user volumes shift from being predominantly humans to predominately machines, and
- the served value elements evolve from simple constructs of today, e.g. Connect (as in CSP services), or 'Connect, compute, store' (as in Cloud services) to much more complex connect compute, store, sense and act value constructs catering to autonomous device swarms of the future.

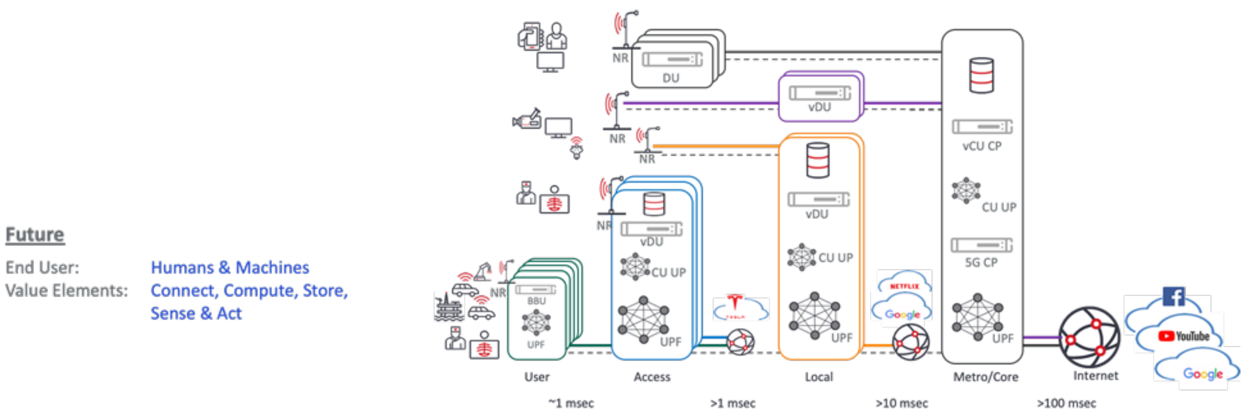

Figure 6: Future State

We believe that pervasive connectivity when complimented with recent advances in Machine Intelligence, Cloud, urLLC communications and the Internet of Things (IoT) will lead to pervasive automation. More specifically, current ML techniques applied to robots and smart devices (and potentially their extension) are a potential tool to enable the mentioned pervasive automation. This will lead to unprecedented immersive experiences provided by swarms of autonomous devices all around us.

These swarms will be enabled by, and be part of, a highly dynamic system formed through an intricate intertwining of connect, compute and store, radically different from today's networks. Within this system, pervasive automation blurs planning, design and operations into a continuum and achieves ongoing optimum operation through appropriate sensing of condition, discerning of meaning, inferring of current/potential deviation from desired operation, deciding on action and acting on these decisions to restore/maintain optimum operation.

<u>Systems Optimization for the 5G era</u>

A 5G connected world is essentially an ecosystem of interconnected intelligence systems. These systems will be vastly complex, intricately intertwined, vastly interactive with order of magnitude larger information exchange than the current; and will evolve rapidly in directions that will be difficult (if not impossible) to plan, design, and operate using current paradigms. This complexity results from the interrelationship, inter-action, and inter-connectivity of different intelligent components within a system (e.g., network nodes, access points, data centers, etc.) and between this system and its dynamic environment.

Handling complexity of such a dynamically evolving arrangement becomes the key challenge, requiring continuous optimization of resources over all timescales and control loops. This brings Plan, Design, and Operations into a seamless continuum. In this continuum, optimal state is maintained by enabling a capable[1] system to (self-)sense the current condition, discern its meaning in the broader context, infer the current/potential deviation from the desired outcome, decide an optimal course of action to best achieve desired outcomes, and finally, act on these decisions on an ongoing basis. An intertwining of heterogenous capable systems, each with appropriate intelligence exhibiting a common control pattern of Sense, Discern, Infer, Decide, and Act, abbreviated as SDIDA hereafter, forms the genesis of self-optimizing fabrics of the 5G era[2]. Figure 7 shows key building blocks of SDIDA.

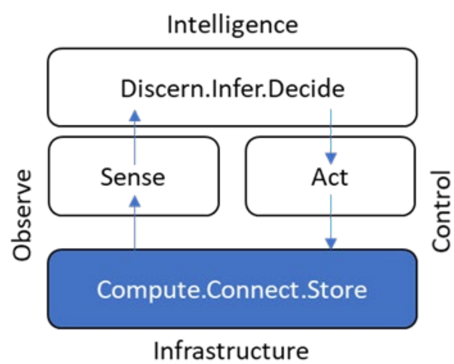

Figure 7: SDIDA Control Pattern

Key building blocks of SDIDA are:

- Discern-Infer-Decide: Represents the 'Intelligence' aspects of the control pattern. They are responsible for collecting data from "Sense," classifying this data (sometimes called feature engineering) and then drawing inferences from the data. The inferences are the basis for insights leading to decisions which are communicated to "Act" as recommendations for execution. The levels of intelligence for a particular embodiment of SDIDA varies vastly depending on the system capabilities, ranging from rule-based simple regression to intricate multi-functional and multi-dimensional reasoning similar to a human brain. Note: in robotics this block is sometimes called "planning" [5].
- Sense: Responsible for collecting telemetry data from the infrastructure and the control and management functions of the entity being optimized.
- Act: Responsible for discerning and executing the logic of a given task. This is accomplished by taking guidance from Discern-Infer-Decide, and by providing instructions for itself or to other

[1] *At one end of the problem continuum a capable system may involve a few transistors, at the other end a capable system may require vast intelligence and insight.*
[2] *https://www.ciena.com/insights/articles/A-Self-Optimizing-Fabric-for-the-5G-era.html*

entities, in turn overseeing their execution (e.g. infrastructure control and management functions of the entity being optimized.)

- Compute-Connect-Store: Represents the infrastructure and respective control and management functions of the entity being optimized (sometimes called "plant" in control theory).

The architecture will also include elements that are common to all building blocks, such as knowledge management processes and components, such as a knowledge base. This is essential to support the intended management of intelligence information to exploit the DIKW process (Data, Information, Knowledge, Wisdom). This is intrinsic to the exploitation of IDN by the elements of SDIDA to retrieve plain data, formalize it as information items, apply different reasoning techniques to obtain as much knowledge as possible, and finally exploit the knowledge to enable SDIDA elements to operate wisely, so demonstrating their wisdom.

The foundational characteristic of SDIDA is the emergent behavior of its entities. Sophisticated behavior is emergent from the intertwining and interacting of simple parts, and the final behavior of the system is not preprogramed or otherwise known ahead of time. SDIDA is recursive, and one instantiation of SDIDA can be controlled by another, more broadly reaching implementation of the same control pattern. SDIDA therefore exhibits a "control of control" pattern.

Provided below is a modular approach to illustrate recursive application of SDIDA to stitch distributed fabrics with various levels of sophistication, leading to self-optimizing fabrics as an nth state ambition. Figure 8 discusses distributed connect fabrics and represents federated SDIDA behaviors.

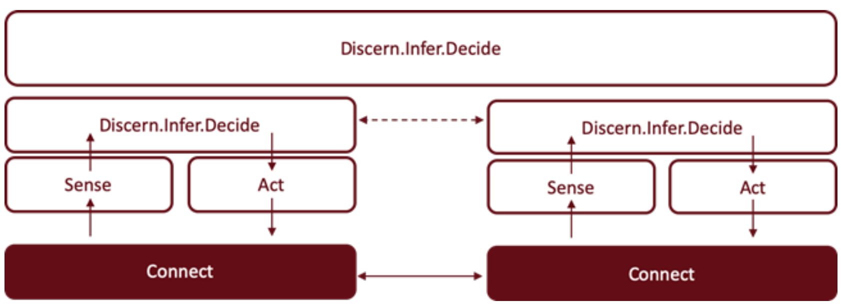

Figure 8: Distributed Connect Fabrics - Federated SDIDA Behavior

Embodiment 1: Distributed connect fabrics; federated SDIDA behaviors:

- Apply SDIDA to static Communication Service Provider Value Element, 'Connect' → resulting construct is Adaptive Connect with close loop automation for singular domain of control
- Stitch multiple Adaptive Connect systems through east-west (peered) and north-south (hierarchical) federation of adjacent SDIDA entities → resulting construct is a Distributed Connect Fabric with federated intelligence

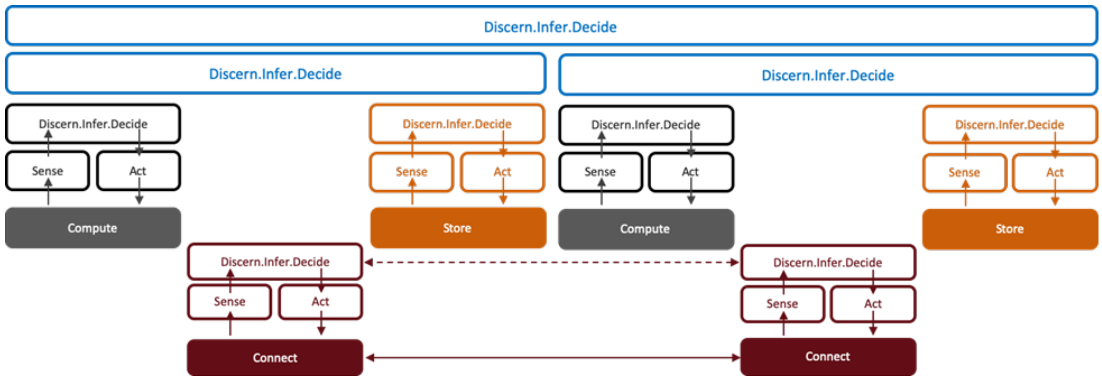

Figure 9: Distributed Cloud Fabrics - Federated SDIDA Behavior

Embodiment 2: Distributed cloud fabrics; federated SDIDA behaviors:

- Extend the Infra Value Elements set from Connect to Connect, Compute & Store, or any combination thereof → resulting construct is Adaptive Cloud for singular domain of control.
- Stitch multiple Adaptive Cloud systems through east-west (peered) and north-south (hierarchical) federation of adjacent SDIDA entities → resulting construct is a Distributed Cloud Fabric with federated Intelligence.

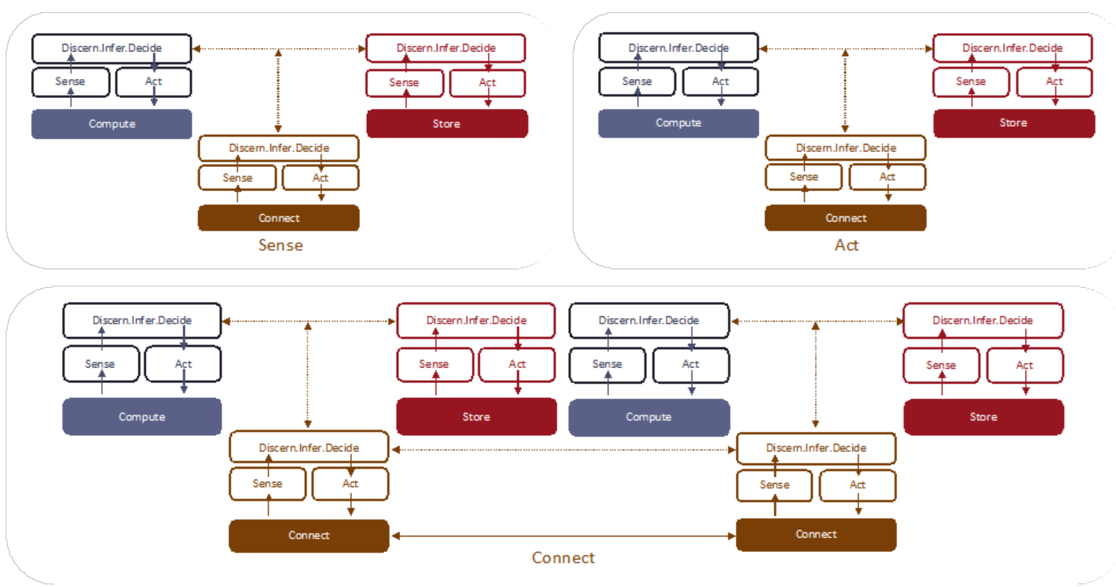

Figure 10: Self-Optimizing Fabrics - Emergent SDIDA Behavior

Embodiment 3: Self-optimizing fabrics; emergent SDIDA behaviors

- In the desired nth state, the overall system will be a fluid federation of distributed intelligences embodied in multiple self-contained intelligence agents. The governing intelligences at various levels (manifested as a dynamic blend of north-south hierarchies or east-west peering, or any other form of adjacent relationships) will be 'emergent' from the interaction of the intelligences in this fluid federation. In the ultimate system, the arrangement of all parts of the system including those of the emergent intelligence will be determined by emergence, i.e., the system will be self-organizing and self-optimizing.

In this nth state all aspects of SDIDA are instantiated, adjusted, and terminated based upon need, policy etc. including SDIDA elements themselves such that a self-optimizing fabric emerges from the environment it controls.

4. Requirements and Technology Gaps

It is important to analyze large systems and study the dependencies among various sub-systems and components within it. It is also important to define the KPIs for various use cases and determine the dependencies and bottlenecks that may affect in achieving these KPIs. It is substantial to test these systems optimization techniques in testbeds.

While the existing standards bodies and fora, namely 3GPP, IETF, ETSI, IUT-T, NGMN and TMF are defining the architecture, framework and protocols, there is a need to develop systems optimization techniques to deliver desired set of KPIs. Hence, IEEE can define a set of optimization rules based on the system interaction and system dependency, while leveraging standards and frameworks being developed in other standards bodies. These rules will need to consider various factors such as machines as the end points instead of humans, closed loop automation using various monitoring and enforcement points, etc. The optimization techniques can depend upon various factors such as traffic variance, control variance and data variance. Machine-learning rules, policies, interfaces, and workflows are missing. Training data set standardization/harmonization is completely open.

We need the feedback from various industry verticals, research labs and service providers who are in the process of deploying these 5G and beyond technologies. Since 5G is an eco-system, it is important to find out the dependence among these components. We will also need to involve various testbeds who are deploying and testing various parts of 5G enablers. We should also get involved with various public-private partnership projects such as NSF funded testbeds, namely PAWR and various other testbeds in other parts of the world. We need the collaboration of SW engineers, data scientists and RF engineers.

5. Roadmap Timeline Chart

Table 1   Working Group Needs, Challenges, and Enablers and Potential Solutions

Name (be brief)	***Current State (2019) (details)***	***3 years (2022) (details)***	***5 years (2024) (details)***	***Future State 10-years (2029) (details)***
Need #1	Dynamic discovery and peering of heterogenous resources	ML-capable entities/fabrics	Architectural evolution for end-to-end autonomic management and control	Dynamic Semantics discovery and negotiation: self-learning protocols to be discovered at the point of attachment

Name (be brief)	*Current State (2019) (details)*	*3 years (2022) (details)*	*5 years (2024) (details)*	*Future State 10-years (2029) (details)*
Challenge(s) for Need 1	lack of entity as well as functionality for performing these tasks	computational complexity, lack of interfaces, lack of data and models	Revolutionary changes in existing architecture	stays in contrasts to today's protocols; requires radical changes in the systems
Possible Solution for Challenge	introduction of a fabric/multiple fabrics into the system	introduction of highly efficient entities/fabrics	self-optimized outer loop	
Need #2	Static protocol and capability negotiation	ML driven dynamic capability discovery and negotiation	Autonomic system behaviors with self-optimized components that leverage any achievements in this area	Dynamic fabric allocation, optimization and monetization with resources contributed by multiple micro data centers
Challenge(s) for Need 2	can be performed locally, but no end-to-end performance guarantee	Need for dynamics ML-driven solutions to guarantee end-to-end performance and adapt to the network dynamics	Definitions of autonomic systems, and abstractions layers for control-loops that close gaps in emerging standards for autonomic networking and autonomic management & control, identification/introduction of self-optimized components, modelling of complex systems,	lack of solutions for enabling and implementing fully autonomous solutions; guarantee of stability
Possible Solution for Challenge	introduction of higher-level fabric to orchestrate/coordinate, additional interfaces/signaling	offline studies and model development and gradual integration	Emergent intelligence solutions	Enhanced emergent intelligence solutions
Need #3	Dynamic capability negotiation	ML driven policy federation across multiple jurisdictions	Autonomic policy negotiation and agreement	Self-determination of federated domains
Challenge(s) for Need 3	no available metric to trigger dynamic capability negotiation	development of policies, enabling federation	Dynamic generation and assessment of policies among multiple domains	Definition of meta-policies to guide the decision boundaries allowed to the network or SON
Possible Solution for Challenge	adaptive triggers	introduction of interfaces between jurisdictions,	Semantic interaction among intelligent agents	Empowering the network with IDN capabilities

Name (be brief)	*Current State (2019) (details)*	*3 years (2022) (details)*	*5 years (2024) (details)*	*Future State 10-years (2029) (details)*
		coordinated multi-agent ML algorithms		
Need #4	ISM, local (private) and national license holder with strict network & spectral resource allocation	ML driven resource federation and optimization	AI powered private network operation and integration with a federated network	Development of new-look internet technology with the federation of private networks
Challenge(s) for Need 4	5G resources available via Network slice from a national licensed operator only.	Definition of realistic problems and practical solutions		
Possible Solution for Challenge	Multi-layer & multimodal resource definition and allocation (e.g. Non-standalone 5G network operation)	Multi-agent, multi-level ML solutions		
Need #5	Need to have a model that can model system dependency and deadlocks	Models that can predict the systems performance based on the schedules and available systems resources	Model should be able to study and detect behavioral properties such as system deadlocks, investigate the anomalies of specific schedules, and then compare various schedules, such as proactive, reactive, and concurrent schedules	Tools that search for application- or context-specific optimizations, such as caching, proactive, or cross-layer techniques
Challenge(s) for Need 5	This model should be scalable and applicable to a large system	This model should be able to design dead-lock free system thus avoiding overoptimization	Automatic generation of schedules for set of operations to provide the desired quality of service with the available resources will help one to use the right set of protocols.	The formalization of key techniques, the models of systems dependencies, and the ability to calculate or predict optimization metrics provide a foundation for the automated discovery and implementation
Possible Solution for Challenge	DEDS (Discrete Event Dynamic	Explore creation of distributed model for an	Thorough analysis of operation protocols that are needed to	Analysis of primitive operations

Name (be brief)	***Current State (2019) (details)***	***3 years (2022) (details)***	***5 years (2024) (details)***	***Future State 10-years (2029) (details)***
	System) -based system model	end-to-end system	execute certain task and map to the model	associated with each protocol
Need #6	Testbed that can be used to test various systems optimization techniques	Federation of Testbeds by connecting various testbeds at various parts of the world	Augment the testbed capabilities to demonstrate various types of applications including augmented reality and other low latency type applications	Integration of some of the advanced techniques and enablers including AI/ML in the testbed.
Challenge(s) for Need 6	Access and availability of a testbed that can be used to demonstrate various proof-of-concepts	Ability to provide scalability to the testbed. Ability to conduct large-scale experiments	Integrate simulation with Testbed. Access to various testing tools, monitoring tools, application generating tools that can simulate a specific part of the testbed	Apply various parameters such as traffic variance, control variance and data variance to demonstrate control and automation in the network
Possible Solution for Challenge (Need 6)	IEEE can serve as facilitator to initiate access to some of these testbeds	Collaborate with public-private partnership testbeds such as PAWR and ENCQOR	Apply the results from the models to the experimental testbed to see the validity.	Evaluate a set of use cases and demonstrate the KPIs in the testbed to see the effectiveness.

6. Standardization Approach

The WG will take the following approach in view of answering the question of "How IEEE work on Future Networks can make contributions to the global landscape of Standards for Autonomic Networking, Autonomic Management and Control (AMC), Cognitive Networking and Self-Management of Networks and Services, while aligning and re-using relevant standards already being developed in various SDOs/Fora in order to avoid re-inventing the wheel":

- Following the approach described in section 1.3, on exploiting emerging standards on autonomics such as ETSI GANA related standards for AMC in diverse network architectures, AMC requirements in the NGMN 5G E2E Architecture Framework, ITU-T standards on AMC in IMT2020, 3GPP related SON standards, IEEE SON related standards, Broadband Forum (BBF) related standards on autonomics in BBF architectures, TM Forum related frameworks on autonomics and autonomous networks, the WG will maintain knowledge on the roadmaps of such standards/frameworks and their applications to 5G and beyond. This helps the WG to identify the standards that address the challenges and problems outlined in the roadmap of this WG. The WG will seek to obtain a clear picture on what the SDOs/Fora are saying are the gaps on standards for autonomics that may be closed by IEEE and any other SDOs/Fora with competence to close those gaps. Then the WG will seek to answer the question of whether IEEE can launch some work on developing the standards required to close any remaining standards gaps. Such an approach helps in enabling adoption and usage of the resultant IEEE developed standards by the industry at large as the IEEE standards would interwork or complement the other autonomics related standards from SDOs/Fora communities outside IEEE. This enables the "Plug-&-Play" of standards from IEEE in the integration with complementary standards from the other SDOs/Fora. The assistance of the Standardization Building Blocks Roadmap WG in sourcing information about standards gaps in autonomics will be sought.

- The WG will use the ROADMAP TIMELINE CHART provided in chapter 5 to derive Architectural Blueprints for autonomics in specific network domains and associated Use Cases of relevance to the aspects outlined by the roadmap, while taking into account any updates to the roadmap as may be necessary. Where there are overlaps with already existing or emerging Architectural Blueprints, Use Cases and Requirements for autonomics (e.g., AMC Requirements and Use Cases specified by NGMN in their 5G E2E Architecture, AMC Requirements and Use Cases specified in ETSI GANA related standards and in ITU, BBF, TMForum, etc.), the WG will seek to avoid re-inventing the wheel. In case of overlaps, the WG will simply adopt the existing or emerging standardized frameworks and enhance their Use Cases with those that derive from the WG' Roadmap timeline chart. The WG will build an understanding of the extent to which autonomics and associated Multi-Layer AI related standards/frameworks from other SDOs/Fora cover the aspects outlined in the Roadmap timeline chart and any WG's resultant derived Architectural Blueprints, Use Cases, and Techniques. As a result, the WG will build a set of Architectural Blueprints and Use Cases that will be used to determine if there are new techniques and new standards required by those Blueprints to address the autonomics Use Cases and Requirements in the Roadmap timeline chart (to be kept updated) that may be developed in IEEE. The WG will keep track of what will be emerging from the autonomics standardization roadmaps in the various SDOs/Fora to align with those roadmaps where necessary. Interactions with the Standardization Building Blocks Roadmap WG will be sustained in order to obtain their insights and collaborate with the WG on standardization matters.
- While the WG will be innovating certain technologies of relevance to 5G and beyond that are not yet existing in the industry (e.g. in the space of quantum computing and other emerging topics), the WG will have the opportunity to create standards on those technologies as IEEE standards. The assistance of the Standardization Building Blocks Roadmap WG in providing insights on what can be considered standardizable in the research results will be sought, and standardization work items shall be derived and launched as a result.

Finally, a few closing remarks on standardization approach:

1. While some of the standards discussed in Section 1.3 and in the appendices, emerging from the various SDOs/Fora identified, are relevant to addressing the challenges and problems outlined in the roadmap of this WG, the WG will keep track of the evolution of such standards in the light of the need to close any standards gaps that derive from the challenges and problems outlined in the roadmap of this WG. Also, it is important to note that the three parallel tracks on standardization approach outlined above will be pursued not only in relation to "5G and beyond" but also in relation to other technologies this WG will consider such as WiFi and other network technologies in which self-optimization and other autonomics functionalities are a fundamental requirement to deploying and operating such technologies.

2. Standardization Imperatives in Summary: Standards provide a foundation to support the development of the kind of innovation that can be more easily accepted and deployed by the industry, in contrast to proprietary solutions that cannot easily integrate and interoperate with other networked solutions. Standards capture tacit best practices and standards set regulatory compliance requirements. Standards support the need to balance agility, openness and security in a fast-moving environment. Standards provide a reliable platform from which solution suppliers can be able to innovate, differentiate and scale up their technology development. They help the industry control essential security and integrate the right level of interoperability. No technical committee or standards organization can single handedly develop all the Standards that are needed. Hence the need for cooperation of IEEE and other SDOs/Fora in order to help innovators and solutions suppliers deliver solutions for the increasingly complex systems such as 5G and other emerging and future technologies in which self-optimization and other autonomics functionalities are a key requirement for leveraging, building, deploying and operating such technologies. Standardization work has several driving forces, such as market demand, use-cases that form clusters and patterns of increasing importance in industry demands at specific timeframes, cost pressures, and stakeholder dynamics within a value chain. Today the challenge is to align and harmonize the multitude of trends of standardization activities so as to capture the synergies among overlapping trends and reduce or consolidate the number of streams working towards partially common objectives.

3. The INGR System Optimization WG in its pursuit of making contributions to the global landscape of Standards for Autonomic Networking, Autonomic Management and Control (AMC), Cognitive Networking and Self-Management of Networks and Services, will continue to align and re-use relevant standards already being developed in various SDOs/Fora in order to avoid re-inventing the wheel. It is thus organizing a workshop(s) for Stakeholder Consultations on the subject to broad base its understanding of the contemporary perspectives of global domain experts from other SDOs/Fora and diverse ecosystem stakeholders.

7. CONTRIBUTORS

(list names of all Working Group contributors to this white paper (Names and Affiliations)

Kaniz Mahdi	VMWare
Meryem Simsek	UC Berkeley
Ashutosh Dutta	Johns Hopkins University Applied Physics Laboratory
Lyndon Ong	Ciena
Aarne Mammela	VTT Technical Research Center of Finland
Jens Voigt	Amdocs
Ranganai Chaparadza	Altran Capgemini/Vodafone
Mohammad Patwary	Birmingham City University
Pedro Martinez-Julia	NICT
Muslim Elkotob	Vodafone
N. Kishor Narang	Narnix Technolabs
Brad Kloza	IEEE Future Networks Initiative
Matthew Borst	IEEE Future Networks Initiative

8. REFERENCES

All cited works to be cited in full and any pictures/graphics/charts used must have proof of permission if not original.

---

[1] C. Bettstetter and C. Prehofer, "Self-organization in communication networks," *IEEE Communications Magazine,* vol. 43, no. 7, pp. 78-85, 2005.

[2] I. Prigogine and I. Stengers, Order out of Chaos, 1984.

[3] N. Davis, A. Dutta, K. Mahdi and M. Simsek, "Solving Complexity with Emergence," in *Proceedings of IEEE 6G Wireless Summit*, 2019.

[4] J. S. Arora and R. T. Marler, "Survey of multi-objective optimization methids for engineering," *Struct. Multidiscipl. Optim.,* vol. 26, no. 6, pp. 369-395, 2004.

[5] J. S. Albus and A. M. Meystel, Engineering of the Mind: An Introductino to the Science of Intelligent Systems, New York: John Wiley & Sons, 2001.

[6] M. Jamshidi, System-of-Systems Engineering — A Definition, Big Island, Hawaii,: IEEE SMC, 2005.

[7] T. Choi, A System of Systems Approach for Global Supply Chain Management in the Big Data Era,, IEEE Engineering Management Review, 2017.

[8] D. DeLaurentis, "Understanding Transportation as a System-of-Systems Design Problem," in *43rd AIAA Aerospace Sciences Meeting*, Reno, Nevada, 2005.

9. ACRONYMS/ABBREVIATIONS

Term	Definition
1G-4G	First Generation to Fourth Generation
3GPP	Third Generation Partnership Project
5G	Fifth Generation
ACK/NAK	Acknowledgment/negative acknowledgment
AF	Autonomic Function
AI	Artificial intelligence
AMC	Autonomic Management and Control
API	Application programming interface
B2B	Business to business
B2C	Business to consumer
BBU	Baseband Unit
BS	Base station
BSS	Business support system
CAPEX	Capital expenditure
CDMA	Code division multiple access
CN	Core network
COTS	Commercial off-the-shelf
CP	Control plane
CSP	Communications Service Provider
C/U	Control plane / User plane
CU	Centralized Unit (vCU = virtualized CU)
D2D	Device to device
DE	Decision Element
DEDS	Discrete Event Dynamic System
DevOps	Development and information technology operations
DFT-s-OFDM	Discrete Fourier transform spread orthogonal frequency division multiplexing
DL	Downlink
DU	Distributed Unit (vDU = virtualized DU)
EAP	Edge automation platform
eMBB	Enhanced mobile broadband
eNB	Evolved node B
ENQCOR	Evolution of Networked Services through a Corridor in Québec and Ontario for Research and Innovation
EPC	Evolved packet core
ETSI	European Telecommunications Standards Institute
FDD	Frequency-division duplex
FDMA	Frequency division multiple access
GANA	Generic Autonomic Networking Architecture
GSMA	GSM (Groupe Speciale Mobile) Association
HIR	Heterogeneous Integration Roadmap
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMS	IP multi-media subsystem
IoT	Internet of things
IP	Internet protocol
IRDS	International Roadmap for Devices and Systems

ISG	Industrial specification group
ISP	Internet service provider
ITS	Intelligent transport system
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KP	Knowledge Plane
KPI	Key performance indicator
LAA	Licensed assisted access
LDPC	Low-density parity-check
LTE	Long-term evolution
M2M	Machine to machine
MAC	Medium access control
MANO	Management and orchestration
MEC	Multi-access edge cloud
MIMO	Multiple input, multiple output
ML	Machine learning
mMTC	Massive machine-type communication
mmWave	Millimeter wave
MR	Merged reality
MVNO	Mobile virtual network operators
NaaS	Network as a service
NF	Network function
NFV	Network function virtualization
NGMN	Next generation mobile networks
NGC	Next generation core
NOMA	Non-orthogonal multiple accesses
NR	New radio
NS	Network slicing
NSA	Non-standalone
OEC	Open edge computing
OFDM	Orthogonal frequency-division multiplexing
OMEC	Open mobile edge cloud
OPEX	Operational expenditure
OPNFV	Open platform network virtualization
OSS	Operational support system
OTT	Over the top
PAWR	Platform for Advanced Wireless Research
PGW	Packet gateway
PHY	Physical layer
PoC	Proof of concept
QoS	Quality of service
RAN	Radio access network
RAT	Radio access technology
RE	Range extension
RSRP	Reference signal received power
SDN	Software defined network
SDO	Standards developing organization or standards development organization
SIM	Subscriber identification module
SLA	Service level agreements
SON	Self-organizing network
TDD	Time-division duplex

TDMA	Time division multiple access
TSDSI	Telecommunications Standards Development Society India
TTI	Transmission time interval
UAV	Unmanned aerial vehicle
UE	User equipment
UL	Uplink
UP	User plane
UPF	User plane function
URLLC	Ultra-low reliability low latency connection
V2I	Vehicle to infrastructure
V2V	Vehicle to vehicle
vEPC	Virtual evolved packet core
VNF	Virtual network function
WRC	World Radiocommunication Conferences
WG	Working group

10. APPENDIX

10.1. APPENDIX A – PREVIOUS WORK ON SELF-ORGANIZING SYSTEMS AND CONTROL

Autonomic Systems: Different definitions are presented in J. O. Kephart and D. M. Chess, "The vision of autonomic computing," *Computer*, vol. 36, no. 1, pp. 41-50, Jan. 2003 and M. Schaefer, J. Vokríek, D. Pinotti, and F. Tango, "Multi-agent traffic simulation for development and validation of autonomic car-to-car systems," in *Autonomic Road Transport Support Systems*, T. L. McCluskey, A. Kotsialos, J. P. Müller, F. Klugl, O. Rana, and R. Schumann, Eds. Cham, Switzerland: Springer, 2016, pp. 165-180. In the former case the autonomic systems are self-organizing computing systems. In the latter case the authors suggest that the system is autonomic if the subsystems are cooperating using a dedicated communication channel just as in distributed systems. The terminology is not unified. Some standards work has been done in ETSI GANA towards standard terminology, see Appendix C.

Feedback or control loops: The feedback loop has various names in the literature such as sense-plan-act paradigm in robotic systems; observe, orient, decide, and act loop (OODA) in combat operations process; decision-making process in situation awareness; cognition or cognitive cycle in cognitive radios; and monitor, analyze, plan, execute, and knowledge loop (MAPE-K) in autonomic computing.

Emergent behavior and hierarchical control: Emergent phenomena may lead to chaotic situations. For example, if there are many connected feedback loops, the system may behave chaotically. A common solution to this problem is to use hierarchy where the upper levels are slow and lower levels are fast.

A good system principle to be followed is **subsidiarity**, which is using multilevel hierarchies of relatively autonomous subsystems, see H. Bossel, *Systems and Models: Complexity, Dynamics, Evolution, Sustainability*. Norderstedt, Germany: Books on Demand, 2007, pp. 46, 47, 214, 280. The author claims that this is the best and most efficient way to organize a hierarchy. Decisions are made where the problems are and thus the system is distributed as mentioned elsewhere in this white paper. Communication between levels is restricted to the essential. The number of hierarchical levels can be reduced to a minimum. Subsystems must therefore be given as much autonomy and responsibility as possible. This improves the effectiveness, efficiency, and efficacy of the system.

This can be used as a general motivation to use hierarchical distributed systems (not for decentralized systems). Note that the subsidiarity principle does not mean completely autonomous subsystems but there is a hierarchy. If a subsystem cannot cope with the situation, the next hierarchy levels takes over. Note also that a completely decentralized autonomous subsystem is optimizing only the subsystems and therefore it will not in general find the global optimum. Furthermore, no technical system can be completely autonomous or self-organizing since technical systems do not have any understanding of semantics or context - therefore human intervention must be always possible.

In such an arrangement, systems form a hierarchy from top down: manually controlled, self-organizing (structure changed autonomously), autonomous (no external control), and automatic (no human control) systems. Learning systems can change their behavior using earlier experience (a memory is needed). Autonomous systems must be learning systems. Automatic systems are systems that do not need any manual control. They include control and adaptive systems and such learning systems that need an external reference signal during operation. So called cognitive systems and artificial intelligence systems are among learning systems, they do not have true intelligence that would imply self-consciousness and understanding of semantics. Note that all technical systems need a goal that is given externally since they do not have free will. The goal may be as simple as a set-point value or more complicated reference signal or reference trajectory or desired performance or desired state.

A good hierarchy is formed in such a way that the **range and resolution** in time, frequency, and space are different at different levels. On top of the hierarchy the range in all these dimensions is long and resolution is low, in the bottom the range is short and resolution us high, see J. S. Albus and A. M. Meystel, *Engineering of Mind: An Introduction to the Science of Intelligent Systems*. New York: John Wiley & Sons, 2001, p. xv. The higher level has the priority to set goals to the next lower level to avoid deadlock situations, see M. Mesarovic, D. Macko, and Y. Takahara, *Theory of Hierarchical, Multilevel Systems*. New York: Academic Press, 1970.

Dependability as a performance measure: In (Avizienis 2004) the authors use the terms functionality, performance, dependability and security, and cost as key performance indicators. Dependability includes the terms availability, reliability, safety, integrity, and maintainability. Security includes the terms availability, confidentiality, and integrity. In the former case availability means readiness for correct service. In the latter case availability means availability for authorized actions only. Note that safety refers to threats from the system to the environment and security refers to threats to the system from the environment, see A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, “Basic concepts and taxonomy of dependable and secure computing,” *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11-33, Jan.-Mar. 2004.

10.2. APPENDIX B – SYSTEM OF SYSTEMS ENGINEERING

System: *A group of interacting, interrelated, or interdependent elements forming a purposeful ‘WHOLE’ of a complexity that requires specific structures and work methods in order to support applications and services relevant to the stakeholders.*

- ***The System is the product of the interactions of its parts, rather than the sum of its parts***.
- ***Systems have properties that none of its parts have (emergent properties)***.
- ***The performance of a system depends on how the parts fit not how they act taken separately***

System of Systems: System of systems is a collection of task-oriented or dedicated systems that pool their resources and capabilities together to create a new, more complex system which offers more functionality and performance than simply the sum of the constituent systems. Currently, systems of systems is a critical research discipline for which frames of reference, thought processes, quantitative analysis, tools, and design methods are incomplete. The methodology for defining, abstracting, modelling, and analysing system of systems problems is typically referred to as “System of Systems Engineering”. [6]

Systems Approach: *A holistic, iterative, discovery process that helps first defining the right problem in complex situations and then in finding elegant, well-designed and working solutions. It incorporates not only engineering, but also logical human and social aspects.*

The System-of-Systems Approach: While the individual systems constituting a system of systems can be very different and operate independently, their interactions typically expose and deliver important emergent properties. These emergent patterns have an evolving nature that stakeholders must recognize, analyze and understand. The system of systems approach does not advocate particular tools, methods or practices; instead, it promotes a new way of thinking for solving grand challenges where the interactions of technology, policy, and economics are the primary drivers. System of systems study is related to the general study of designing, complexity and systems engineering, but also brings to the fore the additional challenge of design.

Systems of systems typically exhibit the behaviors of complex systems, but not all complex problems fall in the realm of systems of systems. Inherent to system of systems problems are several combinations of traits, not all of which are exhibited by every such problem:

1. Operational Independence of Elements
2. Managerial Independence of Elements
3. Evolutionary Development
4. Emergent Behaviour
5. Geographical Distribution of Elements
6. Interdisciplinary Study
7. Heterogeneity of Systems
8. Networks of Systems

The first five traits are known as Maier's criteria for identifying system of systems challenges [7]. The remaining three traits have been proposed from the study of mathematical implications of modelling and analysing system of systems challenges by Dr. Daniel DeLaurentis and his co-researchers at Purdue University [8].

Systems Approach Activities

- Identify and understand the relationships between the potential problems and opportunities in a real-world situation.
- Gain a thorough understanding of the problem and describe a selected problem or opportunity in the context of its wider system and its environment.
- Synthesize viable system solutions to a selected problem or opportunity situation.
- Analyze and trade-off between alternative solutions for a given time/cost/quality version of the problem.
- Measure and provide evidence of correct implementation and integration.
- Deploy, sustain, and apply a solution to help solve the problem (or exploit the opportunity).
- All of the above are considered within a life cycle framework which may need concurrent, recursive and iterative applications of some or all of the systems approach.

of dependable and secure computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11-33, Jan.-Mar. 2004.

10.3. Appendix C – ETSI Generic Autonomic Networking Architecture (GANA)

The ETSI GANA Model (ETSI TS 103 195-2) has defined abstraction layers for designing and implementing multi-layer control-loops (multi-layer autonomics), nesting and hierarchical relationships among control-loops and time-scaling of operations of the control-loops in relation to the hierarchical relationships, interworking of control-loops, centralized control-loops designs/implementation, distributed control-loops designs/implementation, including a framework for addressing "stability of control-loops" and the interactions of fast control-loops with slow control-loops, hierarchical control-loops and nesting, and federations of autonomics components (e.g. federations of GANA Knowledge Plane (KP) Platforms across network segments and network operator domains as discussed in NGMN 5G End-to-End Architecture Framework White Paper (see version 3.0.8) and in the ETSI 5G PoC White Paper No.4 (https://intwiki.etsi.org/images/ETSI_5G_PoC_White_Paper_No_4_v3.1.pdf). ETSI is now working on a Test Framework for Testing GANA Multi-Layer autonomics and associated AI models for cognitive Decision-making-Elements (DEs) that drive specific control-loops at specific

GANA levels. ETSI 5G PoC White Paper No.4 (https://intwiki.etsi.org/images/ETSI_5G_PoC_White_Paper_No_4_v3.1.pdf) also defines three complementary paradigms linked to automation and how they complement each other in what is called "(the 3As)", namely: "Autonomic Management & Control (AMC)"; "Automated Management" and "Autonomous network behavior".

In the ETSI standard ETSI TS 103 195-2, autonomics is defined as the science of control-loops design and implementation (operationalization), nesting and hierarchical relationships among control-loops and time-scaling of operations of the control-loops in relation to the hierarchical relationships, interworking of control-loops, centralized control-loops designs/implementation, distributed control-loops designs/implementation, including frameworks for addressing stability of control-loops and their interactions. The more holistic and generic framework, in terms of abstraction levels sufficient to introduce autonomics in network architectures and their associated management and control architectures, while providing design principles and operation principles for the autonomics, is the newly emerged *ETSI GANA (Generic Autonomic Networking Architecture) Reference Model for Autonomic Networking, Cognitive Networking and Self-Management of Networks and Services*, standardized by ETSI in ETSI TS 103 195-2. Figure 11 presents the snapshot of the ETSI GANA Model.

The Hybrid SON Model is compatible with the ETSI GANA Model as described in ETSI White Paper No.16. The figure below is the snapshot of the ETSI GANA Framework. The ETSI TS 103 195-2 also provides definitions of autonomic behaviours that should be associated with autonomic manager components and their associated control-loops over the Managed Entities (MEs) they are responsible of dynamically (re)-configuring to achieve certain objectives. The autonomic behaviours are dubbed ***Self-* behaviours features*** of a Network Element/Function (NE/NF) or the network as a whole and/or the management and control systems of a network as a whole. The autonomic behaviours include: *self-configuration, self-adaptation, self-optimization, self-monitoring, self-protection, self-defense, self-diagnosis, self-repair/self-healing, self-awareness, etc. As described in ETSI White Paper No.16 and ETSI TS 103 195-2, the ETSI GANA Model resulted from a fusion of a number of leading autonomics efforts/models, including IBM's MAPE Model, 4D architecture, Knowledge Plane for the Internet, and other models, in creating the GANA as a unified reference model.*

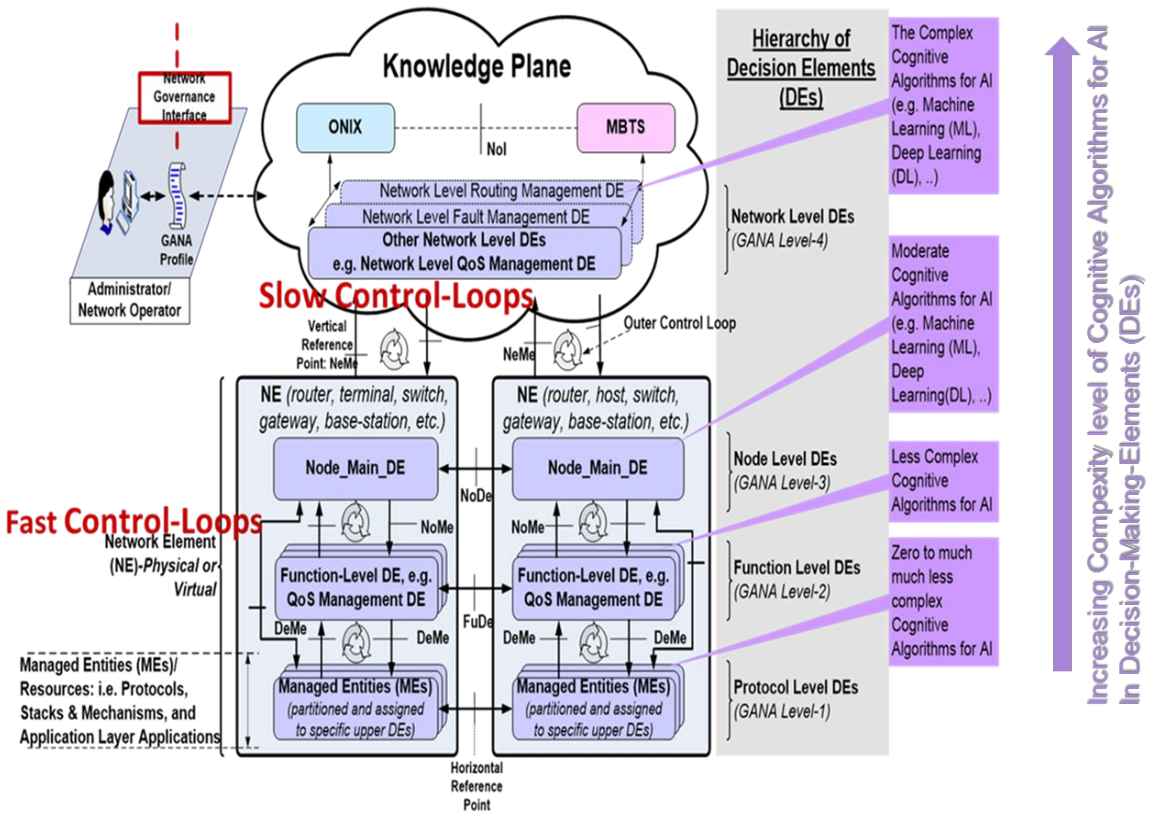

Figure 11: Snapshot of the GANA Reference Model and Autonomics Cognitive Algorithms for Artificial Intelligence (AI), and illustration of the notion of increasingly varying complexity of AI from within an NE/NF (Network Element/Function) up into the Knowledge Plane. Used with permission ETSI TS 103 195-2 V1.1.1 (2018).

The Generic Autonomic Network Architecture (GANA) developed and standardized by ETSI is one such prominent and powerful framework for introducing autonomics in 5G in a standardized way that enables to achieve interoperable multi-layer (multi-level) autonomics, while guiding innovators of autonomics software components (called "autonomic manager components") and the associated algorithms to develop and differentiate themselves by quality of the intelligence exhibited by their respective autonomics software modules. In GANA terms, the "autonomic manager components" are "atomic modules" and are called Decision-making-Elements/Engines (DEs) and referred to as Autonomic Functions (AF). DE algorithms offer for "DE vendor or innovator differentiation" since DE algorithms should not be standardized and should remain IPR (Intellectual Property Rights) for the DE algorithms innovator. Therefore, DE algorithms will continue to be a subject of research and autonomics innovations in the future. As described fully in the ETSI TS 103 195-2, and quoting ETSI 5G PoC White Papers, at the core of any GANA Model is the Functional Block (FB) called the Knowledge Plane (KP) (in reference to Figure 11). The three main building blocks of the KP are:

- GANA Network-Level Decision Elements (DEs), capable of learning and reasoning, and performing decision making planning and actions executions that are meant to realize "macro-level autonomics" at this high level. "Macro-autonomics" (slow control-loops of the KP DEs) is constituted by the KP level DE operations, and is complemented by "Micro-autonomics" (fast control-loops of lower level

DEs) implemented at the NE/NF level. Network Level DEs' scope of input is network wide in implementing "slower control-loops" that perform policy control of lower level GANA DEs (meant for fast control-loops) instantiated in network elements/nodes. As such the Network Level DEs are meant to be designed to operate the outer closed control loops on the basis of network wide views or state as input to the DEs' algorithms and logics for autonomic management and control

- Overlay Network for Information eXchange (ONIX) is a distributed scalable overlay system of federated information servers). The ONIX is useful for enabling auto-discovery of information/resources of an autonomic network via "publish/subscribe/query and find" mechanisms/services it offers. DEs can make use of ONIX to discover information/context and entities (e.g. other DEs) in the network to enhance their decision-making capability.
- Model-Based Translation Service (MTBS), which is an intermediation layer between the GANA KP DEs and the NEs ((Network Elements)—physical or virtual)) for translating technology specific and/or vendors' specific raw data onto a common data model for use by the network level DEs, based on an accepted and shared information/data model.

More detailed descriptions of the GANA Model can be found in ETSI TS 103 195-2 and ETSI White Paper No.16 and other White Papers from the ETSI 5G PoC listed below. The resources also describe how to design and implement GANA Knowledge Plane (KP) Platforms and how a KP Platform can integrate with various management and control systems such as SDN Controllers, OSS/BSS, NFV MANO, Orchestrators, Big-Data Analytics, Ticketing Systems, etc.

With respect to Evolving and Future Networks some of the key merits for technology enhancement by the GANA can be briefly summarized in the following:

- Providing a behavior-centric framework where dynamic operation, behaviour-shaping, workflow-automation, and other key-features for future networking are supported by way of autonomics. Furthermore, this design principle of the ETSI GANA makes it an ideal generic framework for integrating modules and functions in the areas of Artificial Intelligence (AI) and Machine Learning (ML) for autonomics
- Supporting End-to-End (E2E) autonomic management and control of networks and services across network segments and network operator domains as illustrated in the ETSI documents on GANA instantiations onto various reference network architectures and their associated management and control architectures. This means that resource management and control functions (including QoS and QoE mechanisms), mobility, security, and others can be provisioned, managed and adapted dynamically across all tiers of a 5G network or a "beyond 5G network" network, including access, edge, transport, core, data center, backend, etc. network segments. GANA is being applied for autonomic management and control of Network Slices in the ongoing ETSI standardization work, as well as in the ongoing ETSI 5G PoC Project (https://intwiki.etsi.org/index.php?title=Accepted_PoC_proposals).
- Enabling DE innovators to identify GANA abstraction levels for autonomics at which to design DEs, and implement DE autonomic behaviors, as well as enabling Knowledge Planes (KP) federations and DE federations across network segments/domains in order to interconnect peer architectural instances, and enabling closed-loop automation with slow, fast, and nested control loops that follow the GANA principles.

ETSI Technical Committee (TC) INT/AFI WG continues to work with other standards SDOs/Fora such as 3GPP, ITU, BBF, TMForum, NGMN to introduce GANA autonomics in evolving and future networks. There are already a number of documents that involve GANA instantiations onto specific reference network architectures and their associated management and control architectures standardized by specific SDOs/Fora. The following are examples of ETSI GANA instantiations performed by ETSI in collaborations with other SDOs/Fora and are already published and publicly available:

- ETSI GANA autonomics onto BroadBand Forum (BBF) architectures (ETSI TR 103 473 V1.1.2)
- ETSI GANA autonomics onto 3GPP Backhaul and EPC Core Architectures (ETSI TR 103 404)
- ETSI GANA autonomics onto Heterogeneous Wireless Access Technologies using Cognitive Algorithms (ETSI TR 103 626)
- ETSI GANA autonomics in ITU IMT2020 Architectures (ITU-T Y.3324)
- ETSI GANA autonomics onto Ad-Hoc Mesh Networks (ETSI TR 103 495)
- ETSI GANA autonomics in the TMForum ODA (Open Digital Architecture) Architecture

Ongoing further work in ETSI on GANA instantiations include the following (not yet published):

- Implementing Federated GANA Knowledge Planes (KPs) Platforms for E2E Multi-Domain Federated Autonomic Management and Control (AMC) of Network Slices in E2E 5G Architecture; AI-powered Autonomics in 5G Networks
- ETSI GANA Autonomics for IMS (IP Multi-Media Subsystem) Architectures, Closed-Loop Management & Orchestration of IMS Services, Autonomic IMS Service and Security Assurance using Knowledge Planes (KPs) Platforms

ETSI TC INT/AFI WG is also running a 5G PoC (Proof-Of-Concept) Project (a program that is open for any organizations to join) on "*5G Network Slices Creation, Autonomic & Cognitive Management and E2E Orchestration; with Closed-Loop(Autonomic) Service Assurance of Network Slices; using the Smart Insurance IoT Use Cases*" (https://intwiki.etsi.org/index.php?title=Accepted_PoC_proposals). The ETSI 5G PoC project is running various industry Solutions Demos on autonomics and is publishing the results using technical white papers that detail how to implement selected aspects linked to GANA autonomics in 5G network segments and their associated management and control architectures. The following are some of the 5G PoC White Papers that have been published and are available for downloading at https://intwiki.etsi.org/index.php?title=Accepted_PoC_proposals :

- ***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*
- ***White Paper No.2:*** *ONAP Mappings to the ETSI GANA Model; Using ONAP Components to Implement GANA Knowledge Planes and Advancing ONAP for Implementing ETSI GANA Standard's Requirements; and C-SON – ONAP Architecture*
- ***White Paper No.3:*** *Programmable Traffic Monitoring Fabrics that enable On-Demand Monitoring and Feeding of Knowledge into the ETSI GANA Knowledge Plane for Autonomic Service Assurance of 5G Network Slices; and Orchestrated Service Monitoring in NFV/Clouds*

- ***White Paper No.4:*** *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 (KPs)*
- ***White Paper No.6:*** *Generic Framework for Multi-Domain Federated ETSI GANA Knowledge Planes (KPs) for End-to-End Autonomic (Closed-Loop) Security Management & Control for 5G Slices, Networks/Services*

10.4. Appendix D – Disaggregated RAN Architecture

The disaggregation of the RAN currently driven by various open RAN initiatives (e.g., O-RAN) will lead to a variety of new managed functions, entities and applications with clearly defined feature sets within the RAN as well as additional data and policy interfaces between them.

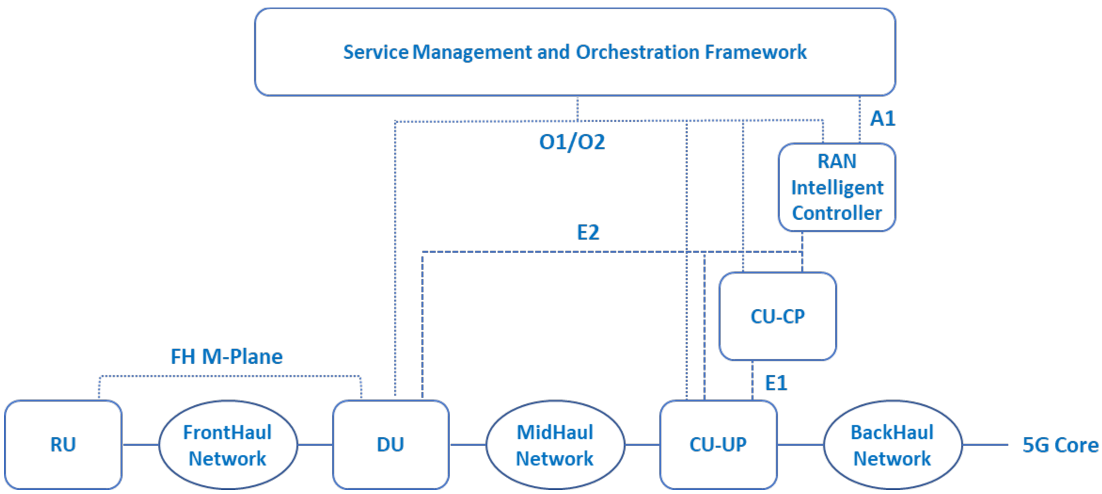

Figure 12: Disaggregated RAN, functional entities and interfaces

The Systems Optimization WG has been formed to explore various approaches to manage complexity of future systems with non-traditional design and operational methodologies. One of the first uses of self-optimizing, self-organizing, or autonomous systems came about in cellular radio systems, with these capabilities specified by NGMN and 3GPP. These systems, however, are based on static policies and are limited in functional scope that addresses 3GPP Multi-RAT for a cell-centric view only. Future generations will include end-to-end non- and near-realtime RAN intelligent proactive control services. The accompanied technology change from linear analytics and convex optimization towards machine-learning based RAN analytics and optimization will allow for UE and UE-group centric services. The deployment of these optimization services will be cloud native.

Another development in RAN technology evolution is the extension of SON (Self-Organizing Network) technology. This was initially deployed by CSPs in a Cell-centric Model in LTE networks. A Hybrid SON Model including both centralized and distributed SON is being deployed by CSPs today and has led to efforts to broaden the scope of control loop-driven management and control beyond the RAN and into other network segments at SDOs such as ETSI, NGMN, 3GPP and ITU-T.

ANTI-TRUST STATEMENT (DO NOT DELETE)

Generally speaking, most of the world prohibits agreements and certain other activities that unreasonably restrain trade. The IEEE 5G Initiative follows the Anti-trust and Competition policy set forth by the IEEE-SA. That policy can be found at https://standards.ieee.org/develop/policies/antitrust.pdf.

This page is intentionally left blank.

FutureNetworks.ieee.org/roadmap