Introduction from the Guest Editors

Nurit Sprecher (Nokia), Vishnu Ram (Consultant),
Csaba Vulkán (Nokia Bell Labs)

5G networks are expected to provide virtually-unlimited-gigabit and ultra-reliable connections to people and objects, when and where it matters, supporting diverse use cases with an extremely demanding range of requirements (in terms of latency, throughput, reliability, coverage and security, cost targets, etc.). However, building a network that supports a broad array and diverse set of new services, all of which can be set up, dynamically reconfigured, scaled and torn down at a moment's notice introduces several challenges. Unprecedented operational agility is required to allow services to be rapidly deployed, dynamically adapted, and continuously and seamlessly assured. The management layer has in its own scope to maintain the overall capabilities of the network, e.g., to resolve any faults, degradations that might have negative implications on the active services. Additionally, it should be able to dynamically manage the available virtual, transport and radio resources in order to maintain the SLA of the established services within the extent of the intrinsic capabilities and resources of the system.

5G network functions can be deployed over virtual infrastructure. The architecture is service based and is designed allowing full programmability and microservice based operation. This yields in a flexible, agile manageability and easy implementation of diverse use-cases with the cost of increased operational and management complexity. Complexity is further increased by the technology complexity (rooted in the 5G architecture and the adoption of IT practices like cloudification, virtualization), the topology complexity (distributed network functions, the 10x densification of RAN sites as well as the hundreds of thousands of network function instances – each with multiple versions) vendor and supplier complexity, dynamic and diverse end user traffic demand that are all increasing the overall cost and the timescale of creating managed services.

The acceleration of 5G network deployments around the world drives the radical change in the way networks and services are created, orchestrated and managed. Evolving towards zero-touch and full end-to-end automation becomes an urgent necessity to manage this complexity, and deliver services with agility and speed, adapt, assure and ensure the economic sustainability of the highly diverse service portfolio.

The ultimate automation target is to enable autonomous networks empowered by advanced cognitive capabilities such as self-configuration, self-monitoring, self-healing, self-optimization and self-protection, without human intervention beyond declaring the high-level business goals.

However, progressing towards this vision requires a new horizontal and vertical end-to-end architecture, which is designed for closed-loop automation and optimized for data-driven analytics and decision-making processes. The ETSI ZSM (Zero-touch network and Service Management) group was formed in December 2017 with the goal to define a future-proof, end-to-end operable framework, solutions and core technologies to enable zero-touch automation of emerging and future networks and services, driven by high-level business policies.

End-to-end automation is a challenging objective and represents the industry's collective target for the technology evolution in years to come. The key components for increasing automation include the expression as well as the translation of a high-level abstract business goals (e.g. intents) into efficient and impactful actions, deployment and interaction with closed loops, real-time data collection, extraction of meaningful insights from data, analytics and intelligence, orchestration and resource control. Many of these aspects still require further research work, learning and sharing of best practices and guidelines in applying them in operational environments of service providers.

The articles of this special issue discuss several aspects related to ZSM, highlight some of the important challenges and further opportunities for research and standardization.

Paul Harvey, Alexandru Tatar, Pierre Imai, Leon Wong and Laurent Bringuier present in their "**Evolutionary Autonomous Networks**" paper, a vision on the future autonomous networking which is based on evolutionary computing capabilities combined with online trial-and-error experimentation to achieve guided emergent behavior. The guidance is provided by technologies such as ontology, to reduce direct human involvement. This approach aims to address the primary challenge of improving the networks' ability to cope with uncertainty, a strong limitation of current, predominantly use case driven approaches to autonomous networking.

The authors introduce the core principles and key technologies, which jointly outline the road towards an autonomous, self-evolving architecture that can automatically adapt to and optimize for any use case. They also highlight necessary standardisation activities that will require industry-wide consensus.

One of the challenging use cases for ZSM is the cross-border vehicular mobility. As the first 5G networks are being deployed across the world, new services enabled by the superior performance of 5G in terms of throughput, latency and reliability are emerging. Connected and Automated Mobility (CAM) services are perhaps among the most demanding applications that 5G networks will have to support. However, CAM operation in multi-operator environments and the inevitable inter-PLMN (Public Land Mobile Network) handover caused by the inherent mobility of CAM services have not been studied in length. Moreover, the multiple domains, multi-vendor components and inherent high mobility in cross-border vehicular environments, introduce multiple challenges in terms of network management and dynamic slicing. In their "**Inter-PLMN Mobility Management Challenges for Supporting Cross-Border Connected and Automated Mobility (CAM) over 5G Networks**" paper, **Konstantinos Trichias, Panagiotis Demestichas and Nikolaos Mitrou** analyse the requirements as well as the challenges of cross-border CAM operation for the five main CAM use cases. Invaluable insights are added regarding the most promising future research directions in the field of cross-border CAM services.

While describing and evaluating the architecture options for CAM, this paper challenges the alignment, mapping of requirements from mobility management to guide further research in real-time applications of closed loops.

Mehdi Bezahaf, Stephen Cassidy, David Hutchison, Daniel King, Nicholas Race, Charalampos Rotsos describe **an autonomic networking framework and procedures** developed as part of the UK Next Generation Converged Digital Infrastructure (NG-CDI) project, in their "**Model-Driven and Business Approach to Autonomic Network Management**" paper. The authors introduce the operators' requirements and propose an autonomic management framework and processes to satisfy a business goal, with recursive capabilities. The business goal derives the overall autonomic decision-making via stages, each of which requires distinct time-horizons, functional logic, and a collaborative loop to link with the previous stage. The authors summarize the current challenges and their plan to use the framework to demonstrate self-adaptation/self-organization function in an operator's network environment.

The ETSI Zero-touch network and Service Management (ZSM) framework introduces standardized components that allow the creation, execution, and governance of multiple **closed loops**, enabling zero-touch management of end-to-end services across multiple management domains. However, the issue of coordination, optimal instantiation and operation of multiple closed loops that support highly dynamic service and operation of 5G systems, is not fully addressed. In their "**intent-driven closed loops for autonomous networks**" paper, **Pedro Henrique Gomes, Magnus Buhrgard, János Harmatos, Swarup Kumar Mohalik, Dinand Roeland and Jörg Niemöller** propose a methodology that introduces requirements to the autonomous management domains and facilitates hierarchical and peer interactions for delegation and escalation of **intents**. The authors propose an extension to the existing management capabilities of the ZSM framework, which allows a conflict-free integration of closed loops by setting optimal goals that each closed loop in the hierarchy needs to account for. They use the network slicing assurance use case to explain the applicability of the proposed methodology. In their summary, the authors highlight the need for Intent standardization in scenarios with multiple intent-driven closed loops, and provide their plan to further investigate the implementation and interoperability aspects of their solution.

Intent can be used to formulate business or service-level abstract goals to the Autonomous Network. The practical implementation of an intent based network requires substantial automation technology embedded in the network. The translation of the intent into efficient and impactful actions in each context requires self-evaluation and self-measurement capabilities. In his "**I^{2}BN: Intelligent Intent Based Networks**" paper, **Péter Szilágyi** examines two key enablers of intent based operation: intelligence and shared context between the intent provider and the network that should autonomously interpret and implement the intent. In intent based networking, this means human intelligence delegated to and implemented in the network, hence the term intelligent intent-based networks or I^{2}BN.

The author investigates the feasibility of automatic creation of intent specific automation pipelines. These pipelines are composed of reusable self-learning closed loop micro-services that are assembled based on their self-declared capabilities. The intent pipeline should be able to contextualize the intent dynamically according to the actual conditions within the network, domain or network function under their control. The paper proposes that the human-network interface should be based on the aggregation of the deployed service and automation capabilities that would eliminate the ambiguity and compatibility gap between the intents defined by the humans and their fulfilment by the network. These concepts are discussed in the context of the ZSM architecture, highlighting the future standardization potential areas (e.g. in ZSM011 on “Intent-driven autonomous networks” and ZSM012 on “Enablers for Artificial Intelligence-based Network and Service Automation”).

Due to the 5G service types and their exigent requirements, the **constant performance monitoring of end-to-end services** becomes crucial to assure the desired requirements. **Nathan Saraiva, Danny Lachos, Christian Esteve Rothenberg and Pedro Henrique Gomes**, propose in their **“End-to-End Service Monitoring for Zero-Touch Networks”** paper a methodology, which is inspired by the ETSI ZSM architecture, for E2E service monitoring. In their approach, a Monitoring Model Generator (MMG) component implements a novel methodology where service deployment and standard information models are used as inputs to automatically generate a high-level monitoring template, Service Monitoring Model (SMM), for a requested **intent-based end-to-end service**. An ontology-based schema is used that is created without including details of the network infrastructure, which facilitates portability and reusability. SMM defines the KPIs and attributes with values and/or ranges for monitoring a specific service based on customer requirements and proposed monitoring ontology. The proposed method extends the ZSM architecture to include service monitoring as a parallel process to service deployment, integrated into the closed loop-based end-to-end service operation.

The authors provide a comprehensive overview of standardization efforts for modelling features of end-to-end service deployment and monitoring. They also propose an end-to-end service monitoring methodology which is based on a domain specific ontology and share a functional validation of the proposed method in a proof of concept implementation which is aligned with ETSI ZSM. The functional validation experiments provide essential guidelines for creating a knowledge base, which is composed of individual instances of SMM, to correlate services with performance metrics and eventually inference new correlations (e.g. SLA violation and its root causes) through knowledge reasoning. The resulting monitoring models are then used to define actual monitoring KPIs and construct management policies in a control loop architecture.

In their paper on "**A model-based approach to multi-domain monitoring data aggregation**", **Antonio Pastor, Diego Lopez, Jose Ordonez-Lucena, Sonia Fernandez, Jesus Folgueira** introduce a **model-based** approach to a **data aggregation framework** relying on standardized data models and telemetry protocols. The multi-domain scenarios considered by the ZSM environments encompass multiple disparate data sources and proper mechanisms for data aggregation, pre-processing and normalization are required to make possible advanced closed-loop management. This heterogeneity increases the difficulty in designing suitable solutions openly applicable and reusable across technologies and deployments. Model-driven data management is based on the idea of applying modelling languages to formally describe data sources, defining their semantics, syntax, structure and constrains on the objects they are associated to.

The ETSI ISG CIM (cross-cutting Context Information Management [8]) considers the exchange of data and metadata across systems a crucial enabler for different applications, and uses the central concept of Context to model relationships between entities according to a semantic graph, providing a theoretical foundation for automated reasoning about the characteristics of the systems they represent and to allow them to better collect information from different origins, filter information from "data lakes" and to create derivative information or decisions. The applicability of the framework has been validated over the INSPIRE5G-Plus testbed addressing a realistic scenario of data aggregation applied to security functions in 5G networks.

Autonomous networks rely on **analytics and intelligence** in order to learn, reason, produce causality analysis, predict, etc. The networks should be able to execute even in conditions of uncertainty and lack of knowledge or under aleatory conditions (randomness/variability). Analytics can learn from collected information, identify patterns and generate models for various optimization and classification tasks. **Artificial Intelligence** (AI) technologies may be used to empower the automation processes. While AI technologies are developing and expanding fast, some gaps must be overcome in order to fully leverage their potential.

AI and Machine Learning (ML) in service providers' networking are still in the early days. We only start to see how AI/ML can be usefully applied. Currently, AI-based solutions are driven by use cases and work in isolation, with limited systematic re-use. The view and use of the AI potential is limited, mainly with Machine Learning/Deep Learning at the descriptive and explanatory levels. AI is still resource hungry and works on big dump data. The integration of AI in network and service automation still requires significant human involvement in all steps of the AI development and operation processes.

Automation is not only about technology. It requires changes in the processes and in the mindset of people. Trust is a key barrier for adoption and building it will be a continuous learning process; as more automation and AI-empowered closed control-loops are deployed and start to operate safely/efficiently, human trust will increase and the requirement for a level of supervision/visibility will diminish. The purpose is to apply transparent, reliable and robust automation.

To increase the comfort level of the human in automation, a level of **supervision** and **transparency** is needed. The ability to explain and trace the machine's decision-making process and the reasoning behind the decisions, are key in supporting it. **Biswadeb Dutta, Andreas Krichel and Marie-Paule Odini** address that challenge in their "**The challenge of zero touch and explainable AI**" paper. The authors describe the challenges to trace back from the inference or prediction events to the original events or symptoms that led to a decision and related actions, and explain the algorithmic approach used to reach the decision. In situations where a series of updated models over a period of time have been applied, even basic information on the model or pipeline, e.g. the version number of the model which produced the inference may not be available. The authors propose an innovative approach which leverages a repository for capturing the taxonomy of all AI/ML models used in the system, as well as architectural components used to annotate events with information regarding the AI/ML model used.

The use of AI in network and service automation will evolve incrementally to the ultimate goals of realizing zero-touch intuitive AI and autonomous operations, leveraging Machine Reasoning (MR) and symbiotic Human-AI interaction.

Enabling zero-touch network and service automation is critical to allow services to be rapidly deployed in 5G and beyond, dynamically adapted, and continuously and seamlessly assured. In addition, it is essential for the constant assurance that the network's performance, coverage and capacity satisfy the service requirements. For us it was an exciting journey to develop and edit this special issue on zero-touch network and service automation and work with the fellow authors. We hope you will find the papers useful, and maybe leverage some of the ideas to help evolving the technology and address the outstanding challenges.

First and foremost, we would like to thank the authors for the insightful and enlightening articles. We would also like to thank the experts who kindly responded to our requests and conducted thorough reviews of the papers and whose constructive comments helped to guarantee the quality of this special issue. The reviewers are in alphabetical order: Laurent Ciavaglia, Stephen Mwanje, Jéferson Nobre, Jing Ping, Vishnu Ram, Nurit Sprecher, Tejas Subramanya, Ishan Vaishnavi, Csaba Vulkán, Mohamed-Faten Zhani.

Finally, we would like to thank Laurent Ciavaglia for his great help throughout the process, his professional advice, recommendations and insightful reviews.

Enjoy reading!