5G innovations for new business opportunities

Salah-Eddine Elayoubi, Jean-Sébastien Bedo, Miltiades Filippou, Anastasius Gavras, Domenico Giustiniano, Paola Iovanna, Antonio Manzalini, Olav Queseth, Theodoros Rokkas, Mike Surridge, et al.

5G PPP

Supported by the European Commission

5G IA

5G
INNOVATIONS FOR NEW BUSINESS OPPORTUNITIES

CONTENT

EXECUTIVE SUMMARY

5G is the next generation mobile network that enables innovation and supports progressive change across all vertical industries and across our society[1]. Through its Radio Access Network (RAN) design and its orchestrated end-to-end architecture, it has the potential to boost innovation and generate economic growth in the European economy. The 5G service models support agility and dynamicity, thereby impacting the granularity, duration and trustworthiness of business relationships. The ability to combine private and public networks and data centres across multiple domains in a secure and controlled way facilitates collaborative business processes. It reshapes the digital business ecosystem with new value chains linking stakeholders from the telecommunications world and the vertical industries in win-win situations. New stakeholders emerge in this evolved ecosystem, for example cloud companies and software houses that profit from the cloudification and virtualization of the infrastructure, and brokers that facilitate sharing of spectrum and trading of connectivity and processing resources. Small and medium-sized enterprises and start-ups are able to embed 5G in their innovative products and services for existing and new customers and markets, leveraging on the Anything as a Service (XaaS) model.

These opportunities are conditioned by the ability of 5G architecture and technologies to deliver the performance levels required for vertical industry stakeholders to engage in the 5G digital business ecosystem. This white paper highlights the technological innovations of the first phase of the 5G Public Private Partnership (5G PPP) and how they contribute to the key performance targets for the 5G service classes: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC). The performance levels ensure an unprecedented experience for end users including high data rates, reduced end-to-end latency, massive connectivity, ultra-reliability and support for very high mobility. The 5G PPP innovations go far beyond what is announced for early 5G deployments. For eMBB service, the integration of mm-wave and frequencies below 6 GHz, along with ultra-dense networks and nomadic nodes, ensure the targeted performance levels with ubiquitous coverage and in high mobility scenarios, in contrast with standalone deployments of mm-wave networks, suitable for fixed usage. The innovations related to the transport network allow also translating the peak throughputs available at the air interface into perceived user experience at affordable deployment cost for operators. In addition to this, the envisioned 5G air interface serves simultaneously all service classes (eMBB, mMTC and URLLC) in a cost effective way, paving the way for new business opportunities with and for verticals.

The 5G concept developed in this paper is not limited to the RAN; it covers the end-to-end path and allows the 5G network to act as a secure, reliable and flexible orchestration platform across multiple domains. The 5G PPP innovations converge towards the vision of 5G as a holistic orchestration platform that integrates networking, computing and storage resources into one programmable and unified infrastructure. The 5G PPP innovations on multi-domain orchestration enable quick end-to-end service deployment and dynamic sharing of networking and processing resources among stakeholders. The 5G security architecture is built on a baseline trust model as a fundamental feature, and provides tools to analyse trust and make it explicit in specific scenarios. The 5G architecture ensures resilience of the network against attacks and its availability during failure incidents. Availability and reliability are achieved by mechanisms such as error recovery, fault detection and fault resolution. These security, reliability and flexibility properties, along with the multi-service air interface, ensure that the 5G network is not just an enhanced air interface as for pre-5G early deployments, announced for the period 2018-2020, but also an open platform for new business opportunities.

The architecture and protocols are designed to adapt to a wide range of deployment scenarios including deep indoor, hot spots, urban areas, rural areas, maritime areas and in an aeronautical context. The 5G concept combines various access technologies, such as cellular, wireless, satellite and wireline, for delivering reliable performance for critical communications and improve area coverage.

Standardization and spectrum regulation are critical elements for avoiding fragmentation of future deployments and increasing efficiency by eliminating redundant options. Spectrum regulation must ensure the early availability of a limited number of frequency bands, which eases the development of the necessary equipment and facilitates faster preparation of tests and trials. As of standardization, 5G PPP projects contribute to 5G standards development by building consensus among European industry, leading to individual and concerted actions towards standardization bodies. In contrast to early announcements of 5G deployments, the 5G results are aligned with the standardization trends in 3GPP, ensuring a global impact of European 5G innovations.

Once the first 5G standards are released and the frequency bands are available, deployments of 5G networks will start, adopting cost efficient upgrade paths building on existing 4G infrastructure. Networking and processing resource sharing strategies between stakeholders can be implemented for delivering the performance targets, e.g. for URLLC use cases, at an affordable cost. This resource sharing/integration is enabled by the multi-domain orchestration advocated by 5G PPP projects and aim at achieving win-win situations for all the stakeholders involved in the 5G service. Regulation must facilitate such flexibility in infrastructure sharing in order to foster the development of the 5G digital business ecosystem.

[1] https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

INTRODUCTION

5G is the next generation mobile network that enables innovation and supports progressive change across all vertical industries and our society[1]. Through Radio Access Network (RAN) design and an orchestrated end-to-end architecture, it has the potential to boost innovation and generate economic growth across all verticals. 5G network deployments and market evolution are subject to the technology achieving the performance targets that accelerate adoption by vertical industries. This white paper highlights the technological innovations developed in the 5G Public Private Partnership (5G PPP) program[2] and how they help reaching the key performance targets for the 5G service classes: enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC)[3]. These performance levels ensure an unprecedented experience for end users including high data rates, reduced end-to-end latency, massive connectivity, ultra-reliability and support for very high mobility, ubiquitously. This white paper shows how the 5G PPP innovations go beyond what is announced for early 5G deployments for the eMBB service class, and how all 5G service classes are delivered over a scalable and cost efficient network. It then explains how 5G technological innovations transform the network into a secure, reliable and flexible orchestration platform across multiple technology and administrative domains. Multi-domain orchestration allows a quick end-to-end service deployment and a dynamic sharing of infrastructure resources among stakeholders, offering new business opportunities and paving the way for new business models.

[1] 5G Vision – The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf
[2] https://5g-ppp.eu/5g-ppp-phase-1-projects/
[3] 5G empowering vertical industries, https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf

15 PPP PHASE 1 GOLDEN NUGGETS

- 5G Spectrum Requirements, Evaluation and Candidate Bands
- 5G System, Functional, Logical and Physical Architectures
- 5G Flexible RAN
- 5G Multi-Service Waveform
- 5G Massive Channel Access
- 5G Flexible Interference Mitigation and RRM
- 5G Performance Evaluation Framework
- Technology Enablers for 5G RAN Platforms (HW & SW)
- 5G Integrated Transport Networks (FH/BH)
- 5G Network Management
- Network Softwarization and Programmability integrating SDN and NFV Technologies
- Flexible and Agile Service Deployment
- E2E Orchestration in Single and Multi-Domains 5G Virtualized Networks
- 5G Networks Security and Integrity
- Programmable Industrial Networks

01 BUSINESS AND STAKEHOLDERS ROLES TRANSFORMATIONS WITH 5G

5G offers new business opportunities on a global level through enhanced performance, flexibility and individualization. Compared to previous generations of mobile networks the changes are more radical. 5G technologies address today's limitations and the future capabilities, such as data rate, end-to-end latency, coverage, softwarization, virtualization, network computing and promise to create hyper-connectivity for delivering unprecedented services in a secure and controlled way. The service levels are able to match the different needs for the benefit of the individual end-customers segments and vertical industries. 5G paves the way for innovative business opportunities for exploiting multiple new Business-to-Consumer (B2C), Business-to-Business (B2B) and Business-to-Government (B2G) business models.

FIGURE 1

Stakeholders around 5G

1.1 NEW VALUE CHAINS IN THE 5G ECOSYSTEM

The 5G service models support agility and dynamicity far beyond what is possible today, thereby impacting the granularity, duration and trustworthiness of business relationships. Network Slicing – a key concept of the 5G architecture – enables such capabilities and allows Network Service Providers (NSP) to develop new offerings using the Anything as a Service (XaaS) model, including IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and NaaS (Network as a Service). The NSPs can flexibly allow the co-existence of multiple tenants on their infrastructure. The tenants, who represent a whole range of different vertical industry stakeholders – called Online Service Providers (OSP) – offer products tailored to the specific needs of their end users. The ability to combine private and public networks and data centres across multiple domains in a secure, controlled and provable way paves the way for collaborative business processes. This enables flexible value chains and value added services in a cost efficient way.

Table 1 provides an illustrative representation of the main stakeholder roles that can schematically be identified, indicating also how their business relationships could potentially evolve as the 5G ecosystem is being developed. In such an ecosystem 5G acts as a catalyst for the development of new business relationships providing opportunities for NSPs (06), a new generation of Communication Service providers (07), Network Infrastructure Manufacturers (02), IT Service Providers (05) and business customers including SMEs. In this new model, partnerships are established across multiple layers of services ranging from infrastructure sharing to network capabilities.

TABLE 1

Stakeholder roles and interactions

	DESCRIPTION	ROLES AND INTERACTIONS
01	IT HW/SW equipment manufacturers	Provide IT HW/SW equipment to IT service providers, Network Infrastructure manufacturers, Network Service Providers and Online Service Providers
02	Network Infrastructure Manufacturers	Provide infrastructure to NSPs (06)
03	Software Network Function Providers	Provide software network functions to NSPs (06), CSPs (07) and OSPs (08)
04	Device Manufacturers	Provide devices to NSPs (06), OSPs (08) and end-customers (10)
05	IT service providers	Provide IT service to NSPs (06), CSPs (07) and OSPs (08)
06	Network Service Providers (NSP)	Provide network services to CSPs (07) and OSPs (08) via Brokers (09), to IT SPs (05), to other NSPs and to end-customers (10)
07	Communication Service Providers (CSP)	Provide communication services to OSPs (8) and end-customers (10).
08	Online Service Providers (OSP)	Provides on-line services to end-customers and can receive content from them. Vertical stakeholders (automotive, energy, factories, health, media) can take this role.
09	Brokers	Intermediary between OSPs and NSPs, and between NSPs in their effort to dynamically establish the most effective solution meeting their needs
10	End-customers	B2C, B2B and B2G

THE ABILITY TO COMBINE PRIVATE AND PUBLIC NETWORKS AND DATA CENTRES ACROSS MULTIPLE DOMAINS IN A SECURE AND CONTROLLED WAY FACILITATES COLLABORATIVE BUSINESS PROCESSES.

1.2 EVOLVED ROLES FOR NETWORK SERVICE PROVIDERS

Main challenges and uncertainties in this changing ecosystem are related to how the NSPs evolve their current business models, to enable the offering of specialized services. Telecom operators are currently facing several dilemmas with respect to business model evolution, multi-stakeholder coordination, alignment of incentives, regulation and competition[4]. A fruitful evolution of the telecom operator oriented business models towards integration of verticals in win-win partnerships is instrumental in bootstrapping and enabling the ecosystem evolution. The major challenge for the NSP is to deliver the needed level of service to a vertical (SSLA: Service and Security Level Agreement), while keeping the control of its own and whole infrastructure (sovereignty).

Changes are induced in the relationships between NSPs, content providers and content delivery providers. The changes enable new service experiences such as immersive media or health services, enabled by orchestrating, controlling, using and monitoring infrastructure resources in an end-to-end coordinated approach. An NSP may negotiate specific wholesale agreements with content providers and content delivery providers for the provision of services allowing greater customer choice and control. In this same context, these agreements can extend to the deployment of in-network content caches, thus enhancing the user experience while mastering network and cache deployment costs.

Revenues for Business-to-Consumer mobile data services in combination with revenues from wholesale relationships will increase when the new services and products are deployed across the value chains. The goal is to create value by detecting new demand for services, enriched by digital platforms, addressing new consumer and business needs.

4 Service Level Awareness and open multi-service internetworking, http://www.networld2020.eu/wp-content/uploads/2016/07/NetWorld2020_WP_Service-Level-Awareness_Final_June-16.pdf

1.3 NEW BUSINESS ROLES AND NEW ACTORS

Vertical industry stakeholders' involvement in the 5G value chain marks the most important change compared to 4G. Stakeholders from vertical industries such as automotive, energy, factories, health, media, public transportation, aeronautics and other sectors, can take the role of OSPs providing services directly to end-customers on top of the infrastructure and connectivity services of NSPs. Manufacturing companies producing vertical industry specific equipment may play the role of device manufacturer.

The introduction of the cloud computing model into the telecom industry enables the emergence of new stakeholders from the IT world (5) into the 5G ecosystem. IT Service Providers and Network Functions Providers can deliver new services such as cyber security services or big data analytics to other stakeholders in the ecosystem such as NSPs, OSPs or directly to end-customers.

The ecosystem and regulatory evolution enables new business role, such as the Broker (9) role that offers services to help its customers be more effective. A broker acts as intermediary between OSPs (including verticals) and NSPs, and between NSPs, in their effort to dynamically establish the most effective solution meeting their needs. Among these brokers we can cite spectrum brokers that facilitate spectrum sharing between NSPs, and connectivity/processing brokers that, e.g., manage marketplaces for trading connectivity and processing resources between NSPs and cloud providers for the purpose of setting up end-to-end services with guaranteed Service Level Agreements (SLAs). Brokers may also act as intermediaries between end-customers and NSPs/OSPs providing services with similar characteristics as for the wholesale market.

To sum up, new business opportunities emerge for telecom/network operators, manufacturers and solution providers as well as for a range of new stakeholders such as OSPs, software houses, brokers, start-ups and SMEs that use 5G for creating innovative products and services for existing and new customers and markets, leveraging on the XaaS model. These opportunities are conditioned by the ability of 5G technologies to provide the targeted performance levels that convince vertical stakeholders and allow the creation of this new dynamic ecosystem around 5G networks.

02 UNPRECEDENTED PERFORMANCE FOR IMPROVED USER EXPERIENCE

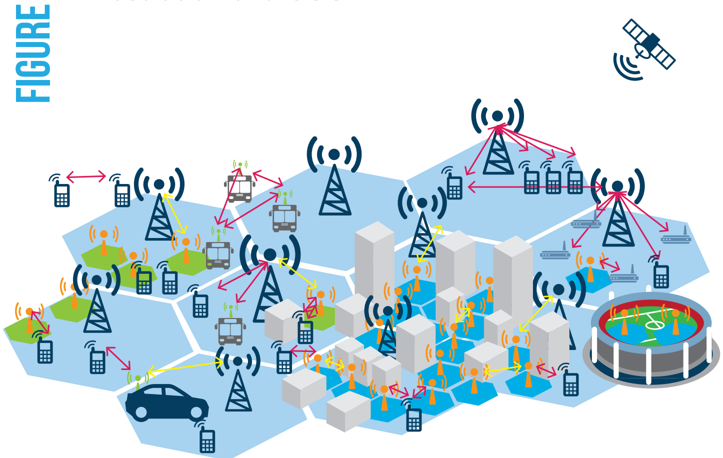
2.1 A FLEXIBLE 5G RAN DESIGN

The promise of 5G starts with the ability to offer an unprecedented experience for end users, in both Business-to-Business and Business-to-Consumer relationships. This includes extremely high data rates, very low latency for devices and support of very high mobility speeds and massive connectivity. This section describes how flexible 5G RAN design, innovations on the physical layer, radio resource management and protocol design, enable reaching these targets.

Due to the diverse requirements of the 5G service classes, the 5G RAN is designed to operate over a wide range of spectrum bands, from 700 MHz to tens of GHz, with diverse characteristics, such as channel bandwidth and propagation conditions. The 5G RAN allows integrating Long-Term Evolution Advanced (LTE-A) technologies, novel 5G radio and WiFi evolutions. Multi-antenna concepts and advanced multiuser detection techniques are integrated at system level and help achieving extremely high data rates across the coverage area. Novel densification strategies lead to the deployment of ultra-dense networks, with fixed small cells and nomadic nodes, such as relays mounted on vehicles. Further capabilities are the native and efficient support of multi-connectivity, Vehicle-to-Anything (V2X), network-controlled Device-to-Device (D2D) and satellite communication. The 5G RAN supports a wide range of physical deployments, from distributed base stations to centralized cloud-RAN deployments or distributed edge clouds. Self-backhauling is an important feature, in which devices act as base stations and self-establish wireless backhaul links to suitable donor base stations [METIS-II]. Support for heterogeneous backhaul technologies is important to maximise 5G availability, resilience and coverage.

FIGURE 2

Illustration of the 5G RAN



INTEGRATION OF SPECTRUM BELOW 6 GHZ WITH MM-WAVE BANDS RESULTS IN AN 11-FOLD CELL EDGE THROUGHPUT IMPROVEMENT COMPARED TO 4G. A 100-FOLD AREA CAPACITY INCREASE CAN BE OBTAINED BY USING SMALL CELLS AND APPLYING AN FLEXIBLE SPECTRUM MANAGEMENT FRAMEWORK.

2.2 EXTREMELY HIGH DATA RATES

The high data-rate requirements for 5G call for a substantially larger amount of spectrum, higher spectral efficiency, and significantly denser deployments of base stations. A particular challenge is the inhomogeneous distribution of traffic over time and space. This requires the network to react quickly and dynamically to fulfil the increased service demand during a time period at a certain region.

These requirements are addressed by a set of key innovations proposed for the 5G Radio Access Network (RAN), covering spectrum usage, flexible Radio Resource Management (RRM), ultra-dense network deployments and highly efficient and low cost transport network.

New spectrum will be available for 5G, including in the millimetre wave (mm-wave) band. The joint usage of high and low spectrum frequencies combines the benefit of higher bandwidth and beam-forming capabilities to increase area capacity using higher frequencies, while maintaining good area coverage capabilities using low frequencies. The flexible and dynamic resource allocation between different radio accesses enables offloading functionality in high resource demand situations. As a result of better resource utilization, the cell-edge user throughput is improved, for example from 15 Mbps to 170 Mbps – an 11-fold improvement compared to baseline LTE [mmMAGIC]. The availability of new spectrum is accompanied by a flexible spectrum management framework that applies new spectrum access schemes in addition to conventional spectrum management methods. For example, License Assisted Access (LAA) and License Shared Access (LSA) can provide more spectrum capacity to end users [Coherent]. A spectrum manager allocates then radio spectrum and energy according to the needs of the traffic, considering the conditions of the surrounding radio environment and the available resources, yielding a 100-fold area capacity increase over current state of the art of LTE small cells [SPEED-5G].

A further innovation that increases data rates is the 5G flexible RRM framework that steers base-stations to use particular channels and spectrum, makes use of dynamic Time Division Duplex (TDD), and allows coordinated operation of access points. Dynamic TDD is the flexible use of spectrum for uplink or downlink based on the instantaneous requirements of burst traffic. In order to contest cross-interference among base stations and among devices, advanced interference coordination ultimately yields an estimated 60 % gain in uplink performance for dynamic TDD compared to baseline, static TDD. Combined with efficient cancelation of very strong cross-interference an additional 50 % gain is obtained [METIS-II]. The flexible RRM framework includes support for fine-grain radio resource allocation and node cooperation across the cell boundaries, improving the performance of coordinated multi-point transmission and reception (CoMP), multi-connectivity and other cooperative functions in the RAN [Coherent]. Aggregating the bandwidth of several network access technologies (cellular, WiFi, satellite and wireline) through traffic splitting/steering techniques is also part of this flexible RRM and helps providing the desired quality of service level throughout the coverage area. These RRM techniques are facilitated by the development of resource pooling techniques (centralized RAN), allowing centralization of RRM decisions. A new architectural paradigm, called Virtual RAN (VRAN) can ensure hardware re-configurability through the utilization of open source platforms and general purpose hardware, thereby extending radio resource allocation flexibility to the processing resources [Flex5GWare].

A further complementary advancement is the deployment of ultra-dense networks, including intelligent small cell nodes and the use of nomadic nodes. Intelligent small cell nodes have the ability to make local decisions on the usage of resources or spectrum bands, improving scalability and reducing signalling loads in ultra-dense environments. The flexible RRM optimises then the deployment of neighbour cells and matches the wireless link requirements [Speed5G]. The 5G infrastructure can integrate complementary nomadic access nodes that increase network capacity and extend the coverage area. For example, by selecting a nomadic node out of 20 randomly distributed nomadic nodes in a macrocell closer to a hotspot, the downlink 10 percentile user throughputs can be improved by around 150% relative to a fixed picocell [METIS-II]. Integration of cellular with satellite solutions ensures the service continuity for certain moving nodes. Note that the increase of data rates through network densification may increase network energy consumption. Therefore it is essential to anticipate the dynamic activation and deactivation of network nodes, such as small cells, to improve the overall energy efficiency of the network. Centralized traffic scheduling and multi-cell coordination schemes, such as dynamic point selection and joint transmission can be used to operate an optimal number of nodes in the network. Relative to no multi-cell coordination, the network power consumption can be reduced by 50% at low user generated traffic [METIS-II].

In analogy to the innovations on the radio interface, the transport network, for backhaul and fronthaul applications, requires much higher capacity. To meet higher bandwidth demand, new functional splitting models are defined, leveraging the distribution of radio functions between radio unit nodes and centralized processing nodes. A single efficient transport network has been designed that is able to support backhaul and fronthaul at the higher data rates at a lower cost [5G Crosshaul]. For example, a 200 MHz frequency band, using 256 QAM (Quadrature Amplitude Modulation), and 8x8 beamforming is capable of generating a peak throughput at the radio splitting interface of tens of Gbit/s, resulting in an aggregate traffic of several hundreds of Gbit/s. This traffic volume is only sustainable through advance optical transmission technologies and a flexible high-speed low-cost transmission using a passive wavelength-division multiplexed network (WDM-PON) was demonstrated [5G-XHaul]. High levels of centralization maximizes the advantages of pooling for transport and processing nodes, while an efficient use of spectrum increases the amount of available bandwidth per user. A trade-off between spectral efficiency and transport optimization resulting from centralization of processing must be found. A novel application for managing energy consumption has been defined, which allows switching transport nodes on or off depending on the activation or deactivation of radio nodes [5G Crosshaul].

While these technologies are tailored for the eMBB service class, they go beyond what is announced for early 5G deployments. The integration of mm-wave and frequencies below 6 GHz, along with nomadic nodes ensures a ubiquitous coverage with high data rates, in contrast to standalone deployments of mm-wave networks. The innovations related to the transport network ensure that the peak throughputs available at the radio interface translate into perceived user experience and affordable deployment cost for operators.

2.3 A VERY LOW END-TO-END LATENCY FOR TIME-CRITICAL SERVICES

Low end-to-end latency is a primary requirement driving the 5G development. Many critical use cases, such as Augmented Reality, Precision Medicine and remote assisted robotic surgery in Health, road safety and autonomous driving in connected vehicles, factory automation, etc., require very low response times in the communication between the respective parties. The delay targets range from tens of milliseconds to 1 millisecond. Network conditions, computing load and congestion induce variability in the end-to-end latency. We show in the following how the latency can be reduced by an optimization of the radio access, the backhaul but also the processing time that incurs in providing the specific service (including the availability of data).

A LATENCY AS LOW AS 1 MS CAN BE OBTAINED ON THE AIR INTERFACE BY APPLYING A NEW FRAME STRUCTURE WITH THE POSSIBILITY OF USING SHORT TTI AND ADVANCED HARQ. MOBILE EDGE COMPUTING REDUCES END-TO-END LATENCY FOR CRITICAL SERVICES.

A low latency air interface

New waveforms that are robust against time-offset do not require signalling associated with time alignment, allowing for reduced latency. A flexible frame design allows multiplexing of Transmission Time Intervals (TTI) with different lengths on the same spectrum resources, reducing latency for short critical messages [FANTASTIC-5G]. Furthermore, advanced Hybrid Automatic Repeat Request (HARQ) schemes have been developed to reduce latency on the air interface based on early detection of packet error without full decoding, or estimation of the number of re-transmissions needed for successful decoding [mmMAGIC]. Multi-mode connectivity, a concept that allows a mobile user to be connected to multiple access nodes at the same time, introduces also spatial diversity allowing information transmission via the shortest path [FANTASTIC-5G].

A further key enabler for reduced latency is to improve control plane procedures and related signalling [METIS-II]. For example, a critical requirement in 5G is that an inactive device with the sudden need to send mission-critical data can quickly access the system, establish a Radio Resource Control (RRC) connection, and send the data. One proposed innovation in this respect is to allow differentiating critical from less critical services through a novel form of preamble use of the Random Access Channel (RACH), which allows increasing the reliability of mission-critical system access at the first attempt by two orders of magnitude, which results in improved system access latency. In addition, a new RRC state is proposed, which allows temporarily inactive devices and the infrastructure to maintain the context information related to a previous RRC connection, for example the security. With this new state, devices can switch about 4-10 times faster from inactive to active state. The exact gains depend on the radio and core network signalling latency. Similar gains are estimated for the downlink. The new state helps to substantially reduce the delay between the paging request to an inactive device and the actual data transmission to the device.

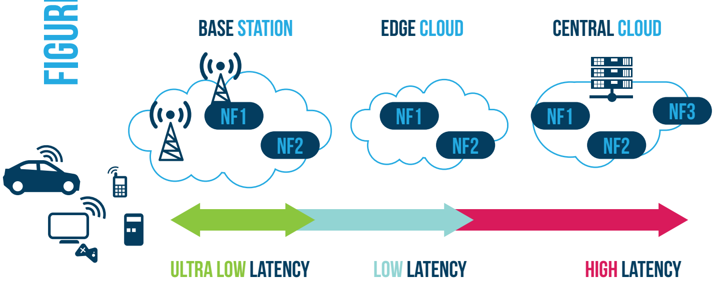
When operating at mm-wave frequencies the objective is to concurrently fulfil high throughput and low latency requirements. mm-wave allows the use of short TTIs, which leads to inherently low latency from a frame structure point of view. Responding to the challenge of processing large amount of data in a short time, high throughput and robust decoders for Low-Density Parity-Check (LDPC) codes to reduce decoding latency are being implemented [mmMAGIC]. The time required for initial access when a user joins the network or performs handover involves a procedure with considerable overhead. Fast beam-alignment schemes that exploit advanced beam codebook design, context information and multi-node coordination have been proposed to reduce the initial access latency [mmMAGIC].

A new architecture that reduces the end-to-end latency

The communication delay caused by the physical distance between the source and the destination is reduced by the introduction of Mobile Edge Computing (MEC) in 5G networks. MEC offers additional processing capacity near the base station for local application level processing. The distance herein refers to the end-to-end network path and rarely refers to the geographic proximity. Decreasing the number of entities along the network path is an established technique to significantly reduce latency. The integration of distributed cloud resources with a cluster of small cells at the network edge assists in achieving low latencies by removing the overhead of backhaul to the core network and enables services at the network edge [SESAME].

FIGURE 3

Intelligent function placement for customized latency



A further enabler for low latency is flexible network architecture, where the various networking functions, such as PHY algorithms, scheduling, HARQ, handover, routing and others, execute in a central or in an edge cloud hardware depending on the specific service requirements. A low-latency service is provided by network functions belonging to both radio access and core network, where the most latency critical functions are moved from the central cloud to the edge cloud, to avoid backhaul latency and unnecessary hops along the path to the central cloud. A service creation request is mapped to network functions that are realised in the instantiation of a suitable network slice configuration taking into account the service requirements. For example, the most critical network functions are moved from the core network into the edge cloud, close to the user. For even more stringent latency requirements, the network functions can be placed on an edge cloud located directly in the base station [NORMA].

A low latency processing in the hardware/ software platforms

In addition to the networking delay between the source and destination, the telecommunication hardware itself introduces latency [Flex5GWare]. The dynamic reconfiguration of hardware and software platforms is being developed to reduce such delays. A hardware-agnostic mechanism guides the reconfiguration of the underlying heterogeneous hardware infrastructure and 'stitches' software and hardware components at runtime, removing repetition of functionality among multiple layers, technologies and elements thereby offering a flexible, reprogrammable and reconfigurable functional composition. A cognitive dynamic hardware and software partitioning contributes to the reduction of the overall latency via choosing configurations that balance latency against other key metrics, such as energy efficiency. This solution considers the latency incurred by all components implemented in either hardware or software, communication between components and external interfaces.

In-network caching as an enabler for low latency communications

Video and augmented reality applications from the media and entertainment vertical industry require low latency coupled with large data rates. A subset of these applications involves videos that are not generated online and can profit from efficient content caching techniques near the end user, reducing the latency associated with content retrieval from remote locations. 5G PPP projects are working towards a flexible content caching solution relying on Network Functions Virtualisation (NFV) and a Software Defined Network (SDN) enabled traffic optimization. In addition extending the cache of the operators to a large virtual cache spanning multiple operator networks is enabled by inter-slice, multi-tenant cache peering technologies and supplements the advantage of caching in the edge cloud. A related innovation reduces directly the latency introduced by network nodes using a special optimization of IPv6 addresses to speed-up the routing process. In a hierarchically structured network the IPv6 address prefix is evaluated gradually by the routers, eliminating the need for a routing lookup-table, which is the most time expensive operation in IP-routing. Cache efficiency is increased by mechanisms for content pre-fetching at the edge nodes including base stations, small cells, nomadic nodes and even User equipment (UE), using context-aware information. Content prefetching and cache updates may be performed using multicast/broadcast capabilities of terrestrial networks and satellites, alleviating the signalling and traffic loads in the middle and last mile [Charisma].

The latency-reduction techniques are complementary and may be used to meet the requirements of the targeted service. For example, when considering tele-operation of moving robots with haptic feedback, the latency reduction techniques are applied for handling control and haptic feedback signals on the radio interface and along the path between the two end points, while fog computing near the robots is used for image processing and local control algorithms.

Many of the technologies used to increase the data rate, contribute to the reduction of latency. Flexible RRM allows multiplexing time-critical services with other services, ultra-dense networks ensure the availability of a close access point and RAT integration and multi-connectivity allow reaching the destination by the shortest path. However, centralization of processing – for interference management and resource pooling – may be conflicting with the need of performing application-level processing near the end user. A hybrid and flexible architecture where processing is centralized for some services and distributed for others is needed.

2.4 A MASSIVE CONNECTIVITY OF DEVICES

The commercialization and deployment of 5G systems is driven by the need to support very high connection densities to make the Internet of Things serviceable. Connection density is defined as the average number of simultaneously active connections that can be supported in a given area, measured in connections per square kilometre. Example use cases include crowded spaces, i.e. stadiums or concert venues, as well as massive MTC in cities, agriculture or factories, etc., where sensors, actuators, and controllers are wirelessly inter-connected. Massive connectivity is supported by new air interfaces that should optimise the available radio and infrastructure resources, spanning areas from protocol enhancements and radio resource management to waveform design [mmMAGIC, FANTASTIC-5G, METIS-II].

One of the enablers for massive connectivity is the handling of the transition of UE modes. Transitions between idle mode and connected mode must be simplified or even avoided. In this direction, connection-less transmission of small packets from UEs once registered and authenticated in the network reduce significantly the required number of signalling messages. In this case, the (small) packet must comprise both source and destination addresses. A component contributing to the reduction of signalling upon connection establishment is the addition of new Radio Resource Control (RRC) states such as the RRC extant state

[FANTASTIC-5G, METIS-II], which is a hybrid state between RRC idle and RRC connected states. From RRC idle it inherits its behaviour, i.e. UE controlled mobility and from RRC connected it retains most of the UE specific access stratum information.

A further innovation is related to the channel access for sensors. Classical channel access protocols comprise two stages: the access notification stage and the data delivery stage. This leaves room for feedback and resource allocation to the UE from the base station, for instance, related to power control and timing alignment. New, "one-stage" access protocols are being developed, in which access notification and data delivery are performed in a single transaction by means of one or more consecutive packets or in a single transmission thereby reducing signalling overhead for short messages [FANTASTIC-5G]. Although initially designed to meet very low latency requirements, the "one stage" protocol is also of interest for longer latency channels like satellite links since it minimises the handshakes.

Additionally, new random access schemes are proposed, where, each device is allowed to contend with a predefined sequence of preambles over multiple Physical Random Access Channels (PRACHs), denoted as the device's signature [FANTASTIC-5G]. This signature is constructed from information unique to each device, such as the device's identity. The proposed schemes result in a significant reduction of message exchanges in the access protocol and can be complemented by appropriate collision resolution techniques, exploiting sparsity properties.

A new waveform design is proposed for asynchronous small packet transmissions in the uplink [FANTASTIC-5G]. Because of the superior spectral properties of certain waveforms, the need for tight temporal synchronization of users can be relaxed. This allows compressing or even avoiding broadcast messages, thus leading to energy and radio resource savings.

Finally, interoperability between the aforementioned innovations supporting the mMTC service class and other solutions for data rate increase or latency reduction is ensured by considering a flexible frame structure, allowing multiplexing of short and long TTIs and coexistence of different waveforms on the same frequency band [FANTASTIC-5G].

2.5 A HIGH PERFORMANCE IN HIGH MOBILITY SCENARIOS

Over the past decades, digital communication technologies introduced significant advances in mobility support, such as seamless handover between cells and efficient coverage of highways and railways. Challenges remain for very high mobility scenarios and for supporting applications such as road safety, assisted driving, autonomous driving as well as other business and infotainment applications.

NEW WAVEFORM DESIGNS ALLOW RESISTANCE TO DOPPLER EFFECT AND ENABLE PERFECT TRANSMISSION FOR VEHICLE SPEEDS UP TO 600 KM/H.

Robustness to Doppler Effect – a must have for high mobility scenarios

Fast moving mobile nodes suffer from the Doppler shift and spread. The latter leads to severe inter-carrier interference in 4G and the network access for Intelligent Transport Systems operating at 5GHz (ITS-G5), because these systems rely on the orthogonality property of subcarriers in the orthogonal frequency-division multiplexing scheme. This interference induces reception errors and imposes retransmissions that lead to increased latency and require the use of robust modulation and coding schemes with reduced spectral efficiency. The development of new waveforms that provide better spectral containment of the signal power, reduce the effects of inter-carrier interference induced by the Doppler effect [FANTASTIC-5G]. 5G is being designed with a flexible configuration of radio resources in the time and frequency dimensions, allowing for larger subcarrier spacing in sub-bands, further alleviating the impact of inter-carrier interference. Initial demonstrations [FANTASTIC-5G] provide evidence of the high potential of the proposed solution. In the case of an uplink transmission of an image at a vehicle speed of 600 km/h and using the proposed new waveforms the image was perfectly transmitted, while using current 4G technology, inter-carrier interference caused signal corruption that results in a failed transmission. These results confirm that new waveforms can improve transmission quality such that highly efficient V2X (vehicle-to-everything) communication is enabled.

Seamless handover and multi-connectivity – zero interruption time and increased diversity

High mobility induces frequent network attachment procedures that may cause annoying service interruptions. High capacity backhaul and the concept of synchronization of base stations enable synchronous and random access-less handovers. Base stations agree on the time a handover will take place and the mobile user receives an un-interrupted service that is seamlessly transferred from the source to the target cell at the agreed handover time without a new network attachment procedure [FANTASTIC-5G].

The generalisation of the multi-mode connectivity concept introduced above enables simultaneous connectivity of a vehicle to several base stations at the infrastructure as well as to other vehicles, allowing for robust and seamless handover [FANTASTIC-5G].

The Central Controller and Coordinator (C3) is a logical entity in charge of centralised network-wide or large area-wide control and coordination among entities in the Radio Access Network (RAN) – possibly using different Radio Access Technologies (RATs) – based on centralised network view [Coherent]. It facilitates the handover procedure by programmatically acting reactively or proactively on user mobility events. The proposed seamless handover, multi-mode connectivity and UE-relaying can easily be implemented and managed by the C3. The handover decision is based on knowledge retrieved by the network graphs and hence is more efficient. The centralized network view allows considering several metrics in the handover decision, such as received signal strength, interference level, network load and vehicle speed.

2.6 A UE POSITIONING ACCURACY IN THE RANGE OF SUB-METER

The new RAN design allows achieving a new kind of performance indicator, that is UE positioning. A positioning accuracy in the order of sub-meter is critical for a plethora of Location-Based Services (LBSs), as input for data analytics, to improve public safety in emergency scenarios, to introduce services like collision avoidance in autonomous vehicles and to create new customized experiences and services for the end user. The target sub-meter range is clearly beyond what can be achieved with existing systems. 5G RAN has a few key enablers that can be seized by positioning systems, including in indoor and urban canyon areas:

- The high density of access nodes leads to a high probability of Line-of-Sight (LoS) conditions between access nodes and the UE.
- The large channel bandwidth available in 5G networks enables increased positioning accuracy for methods based on time-of-flight (ToF) distance measurements [Flex5GWare].
- The availability of multiple antennas allows the use of beamforming techniques for positioning for Angle-of-Arrival (AoA)-based techniques, besides minimizing the multipath propagation effects for ToF-based ranging methods [5G-XHaul].

The envisioned integration of LTE-A evolution, with novel 5G radio, WiFi evolution and other technologies such as Global Navigation Satellite System (GNSS) positioning is a key to achieve better positioning accuracy.

03 5G NETWORK AS A SECURE, RELIABLE AND FLEXIBLE ORCHESTRATION PLATFORM

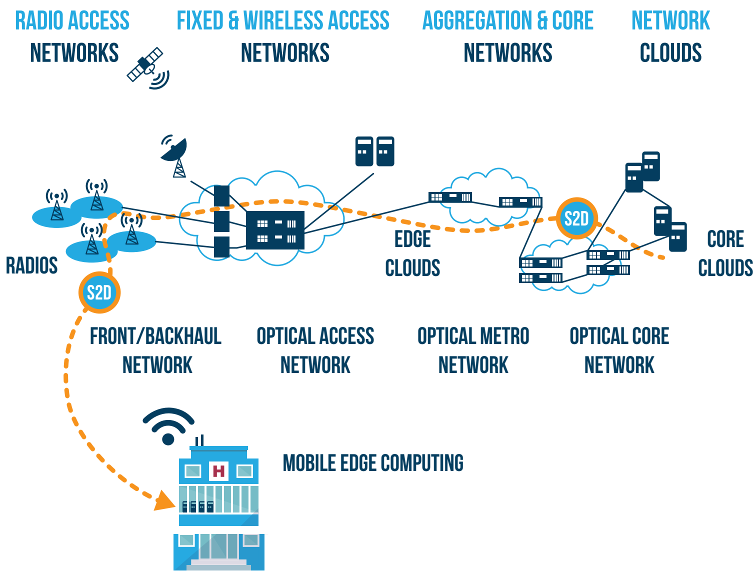
3.1 A FLEXIBLE ARCHITECTURE THAT INTEGRATES NATIVELY NETWORKING, COMPUTING AND STORAGE RESOURCES

5G is a holistic orchestration platform that integrates networking, computing and storage resources into one programmable and unified infrastructure. This vision requires a flexible multi-tenant architecture where computing resources are distributed within the network including sites of the vertical industry stakeholders, within the base stations, in edge clouds at central offices, in regional and central clouds, and managed by different stakeholders. The new architectural paradigm will need to support heterogeneous hardware resources in the same holistic vision: custom ASICs (Application-Specific Integrated Circuit) are still the way to implement fast network processing, and heterogeneous components such as FPGAs (Field-Programmable Gate Array), GPUs (Graphics Processing Unit) can be added to commodity hardware machines to dramatically increase the performance of certain workloads. Therefore, the Virtual Network Function (VNF) concept defined in current NFV architecture should be extended to support a more granular decomposition of the functionality and the mapping into heterogeneous hardware execution platforms.

The extremely high data rate capabilities combined with the availability of storage and processing resources in the edge nodes allows local processing and data analytics close to the users or within the vertical industry stakeholder premises, offering advantages such as low latency, security and confidentiality. Processing capacity in the edge nodes allows software to be transferred and executed near the data (Software to Data, S2D).

FIGURE 4

A flexible architecture integrating networking, processing and storage resources.



A NEW 5G SECURITY ARCHITECTURE INCLUDES A BASELINE TRUST MODEL AS A FUNDAMENTAL FEATURE AND ADDRESSES ALSO SECURITY ON AN END-TO-END FASHION, TO SUPPORT APPLICATIONS THAT REQUIRE COORDINATION ACROSS MULTIPLE DOMAINS.

3.2 A SECURE AND TRUSTWORTHY NETWORK

The security architectures for current 3G/4G networks are defined by 3GPP[5]. However, there are few main drivers that are pushing for a new security architecture for 5G:

- Growing threat levels and increasing abuse of peripheral devices for attacks on back ends and entire networks (for example through DDOS attacks);
- Performance levels for supporting safety and security of critical services in different vertical domains such as eHealth, transportation and industrial automation;
- Introduction of new technologies such as virtualization, edge-computing, fog-computing, in network data and video caching and purpose-specific hardware;
- Partial lack of coverage of management aspects;
- Absence of an explicit and complete trust model for 3G and 4G networks; and
- New business models involving more complex trust relationships enabled by open access via network slicing.

Trusted and Trustworthy 5G Architecture

Without a well-defined trust and governance model, it is unclear which stakeholders have what responsibilities and liabilities in the new business ecosystem. The current ad-hoc approach works well for a small number of network operators. The proliferation of operators is already causing concerns, making network infrastructure more open and posing risks such as impersonation on signalling interchange networks. In 5G networks these problems will become more significant, as will the possibility that security issues are not addressed by any stakeholder, creating opportunities for attacks.

The security architecture being developed in 5G PPP will extend and influence the 3GPP security standards and architecture to capture virtualization and network slicing aspects as well as to include a baseline trust model as a fundamental feature [5G-ENSURE]. While some stakeholders and trust relationships exist in every 5G application, there is no one trust model to fit all situations. Trust and trustworthiness must be geared to the needs of vertical

[5] TS 33.401, 3GPP System Architecture Evolution (SAE); Security architecture.

applications, taking account of the business stakeholders and relationships involved, and the level of risk for each stakeholder from network-related security threats. In the future, vertical industry consortia must agree trust models to meet business and regulatory requirements in a range of critical and non-critical applications including scenarios that cross borders or span multiple physical infrastructure domains.

To support this security architecture, a security management framework that relies on autonomic network management is proposed. Autonomic management solutions leverage insights from real-time analytics, and actuation on network resources in real-time to minimize or prevent the effects of detected threats [CHARISMA].

Multi-Domain and Multi-Layer Security

5G networks rely on network virtualization, implemented by VNFs, forming network slices executed on a shared infrastructure. The 5G security architecture must follow the design principles of the overall 5G architecture, and needs to be logical rather than physical. Slicing must isolate resources and data on shared infrastructure.

Thereby, 5G security must address threats in an end-to-end fashion, to support applications that require coordination across multiple domains. This requires monitoring security in a cross-domain fashion, between physical domains and layered virtual domains. A proposed solution is to use a hierarchical management architecture matched to the trust network, allowing monitoring and control between mutually trusting and trustworthy stakeholders [Selfnet].

Security as a service

5G networks support new business models, enhanced connectivity services and enriched network functionality based on a combination of network operator and vertical industry stakeholder assets and capabilities. In many new business models, the role of a virtual network operator may be fulfilled by a vertical industry-focused organisation such as a manufacturer or a health care provider. Such organisations may not have the capability of managing network security, and may not want to invest in acquiring expertise in areas outside their core business. Security services are therefore needed in conjunction with virtualised network provisioning services to support virtual network operators manage their networks.

Key requirements for security services have been identified and enablers are being developed to meet these requirements [5G-ENSURE]:

- Trust enablers to provide users with possibly certified information about trust dependencies and trustworthiness of stakeholders and technology components;
- Enablers to support authentication, access control and accountability, such as group authentication of Internet of Things (IoT) devices;
- Privacy enablers to improve subscriber identity protection, during network connection and authentication procedures;
- Security monitoring enablers to detect security breaches including malicious or compromised devices using network function and traffic monitoring and analysis;
- Network management and virtualization enablers such as platform integrity attestation.

Such security enablers must be supplied as commodity components and must be able to handle the scalability and diversity of virtual networks needed by vertical industry.

3.3 A RELIABLE AND RESILIENT NETWORK

Availability is related to the service coverage and is defined as the probability that a service request is accepted with the target Quality of Service (QoS). Reliability is the capability of the system to offer a continuous and consistent service quality while operating in dynamic conditions. Quality is associated with performance indicators that depend on the targeted service, such as cell edge data rate for eMBB and latency and packet error rate for critical IoT. In a reliable network the target performance indicator must be met using mechanisms that are intrinsically unlikely to fail, for example redundancy or error correction mechanisms.

Network resilience refers to the ability of a network to recover from harm caused by an event or situation that degrades network QoS. Reliability measures help preventing some types of threats, while resilience measures help recovering from threats that could not be prevented. Network security measures may also be considered in this way, given that QoS in a 5G network also covers confidentiality and integrity characteristics as well as availability and performance.

The use of 5G networks in safety and security critical applications means that threats may have a far higher impact. Measures to prevent or mitigate such threats are of paramount importance, to ensure the effects do not adversely impact critical applications. Measures to ensure a reliable and resilient network for critical applications have to cover networking aspects, such as ensuring a high reliability and availability of the connection [FANTASTIC-5G], and security aspects, such as ensuring isolation between network slices [CHARISMA].

Availability and reliability on the air interface

On the radio interface, reliability and availability must be ensured by means of ubiquitous coverage and error free transmission. A set of innovations on the radio interface is proposed that ensures meeting this target. Coverage extension is ensured by a new control channel design, exploiting beamforming for control channels. To support low end devices in remote areas, asymmetric link operation is proposed in which cell edge devices can use long transmissions for an appropriate signal decoding. To enable robust communication with low latency, advanced error correction and recovery mechanisms are proposed. On a system level, 5G utilises multi-connectivity, in which messages can be sent simultaneously over several radio links allowing spatial diversity and possibly combined at the receiver for a robust decoding [FANTASTIC-5G]. Satellite-terrestrial network integration provides a relevant contribution to the overall set of deployment options and helps increasing network availability. Combining several technologies at the radio interface and at the backhaul may be needed to improve the service reliability especially to support mission critical applications.

Reliability in virtualized networks

The reliability of end-to-end services is hard to assess in an NFV environment in which network functions are dynamically deployed and share the same hardware infrastructure during their execution. This is due to the multi-layer dependencies introduced by the usage of a common infrastructure and virtualization platforms by multiple network functions which splits the management concerns in multiple and transparent to each other layers. Although managed separately all these layers influence the reliability of the end-to-end service. The addition of the virtualization layer between the hardware and the software that acts as a broker for available resources results in a less stable infrastructure compared to the previously used physical network equipment infrastructure.

To fully benefit from system dynamicity and elasticity in deployments over virtualised infrastructures, a reliability framework – based on machine learning – is proposed that works in two main directions [Cognet]:

- Anomaly detection of patterns in data that do not conform to expected behaviour and enablement of system adaptation to unforeseen conditions. Anomaly detection employs a semi-supervised learning approach, which constructs a model of normal behaviour from a training dataset representing normal operation. The deviations from the normal behaviour are used to detect potential anomalies.
- Fault detection, isolation and resolution of network malfunctions. This function correlates the number of detected incidents and identifies the fault for the observed malfunctions. The fault removal actions defined in the policy specification resolve the identified errors and failures.

Resilience against security threats

Malicious attackers and associated threats are inherently present in any system designed for mass use. We cannot assume that all subscribers are trustworthy – some try to cheat the system or have other motives to degrade it. The network must deploy measures to detect and prevent security intrusions fast, so that the number of intruders in the network remains small. Furthermore it must be intrusion tolerant, which means that it should degrade gracefully rather than abruptly or catastrophically in the presence of intrusions.

Isolation between network slices is an important property in delivering resilience and intrusion tolerance. The proposed security management solution manages interference across multiple slices and allows preventing tenant-on-tenant intrusions with a high level of assurance [Charisma]. Resilience to security threats may involve security measures that could degrade QoS, for example by increasing latency or dropping packets. The objective is achieving and maintaining 5G performance while assuring security. For cases that the primary attack cannot be prevented the best solution is to predict cascades of secondary effects caused by security threats [5G-ENSURE], in order to identify measures that could prevent the propagation of adverse effects.

Service continuity during disaster scenarios

CContinuity of service during hazard and emergency events is crucial. A number of techniques are being explored, among which UE-relaying and satellite networks. Based on their native multicast/broadcast capabilities, these latter can offer solutions to "thundering herd events" where huge number of data and video requests need to be served by 5G networks.

As of UE-relaying, envisaged by 3GPP as a key technique for supporting coverage extension for public safety applications, it can profit from the Central Controller and Coordinator introduced above. The C3 can improve mobility and network management for UEs attached to a UE-Relay, in areas with poor or no network coverage, such as tunnels, caves, valleys; in-door, such as in buildings, basements; and in emergency situations such as life-saving disaster relief missions [Coherent]. Depending on the technology properties, service requirements, mobility information and type of environment, the C3 can decide how to improve mobility management, how to perform relay selection and how to configure the system for coverage extension in out-of-coverage situations. The C3 configures different network entities for relay selection, coverage extension and handover. This reconfiguration enables providing different service priorities to relay firemen and policemen under network out-of-coverage situations in both urban and rural environments.

These innovations drastically enhance reliability and increase resilience, although in a wireless context performance degradation may be unavoidable. The strict safety requirements of critical applications, imposed by legal constraints, motivate the definition of graceful degradation options for 5G performance indicators such as data rate, latency, reliability, security and trustworthiness, in order to survive the worst case scenarios.

3.4
A QUICK END-TO-END SERVICE DEPLOYMENT

Vertical industry applications require customised access for different stakeholders. Providing customized access to different stakeholders is labour- and time-intensive during installation and commissioning as well as during operations and maintenance (O&M) activities [VirtuWind). Programmable networks and multi-tenant capabilities in 5G ensure fast deployment and new services. This includes the ability to create, sell and provision composite services in multi-domain environments; technology domains (intra-operator) as well as administrative domains (inter-operator). The 5G PPP target time in reaching a complete service deployment is less than 90 minutes.

The following sections present approaches for reducing the service deployment time, starting from a multi-domain orchestrator that enables service creation over multiple administrative and technology domains, before detailing intra-operator orchestration and management tools.

MULTI-DOMAIN ORCHESTRATION ALLOWS A DRASTIC DECREASE OF THE SERVICE DEPLOYMENT TIME. PROVISIONING TIME FOR A NEW ANTENNA IN THE EDGE UP TO ITS OPERATIONAL STAGE CAN BE REDUCED FROM 120 HOURS TO 90 MINUTES. THE SETUP TIME FOR A SERVICE PUBLISHED IN A MARKETPLACE, SPANNING ACROSS MULTIPLE ADMINISTRATIVE DOMAINS, CAN BE AS LOW AS FEW SECONDS.

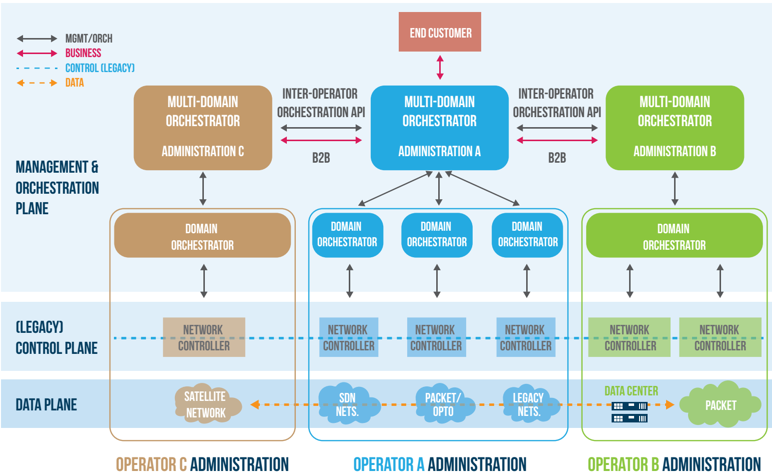
A Multi-domain Orchestrator for reduced end-to-end service deployment time

The enablement of cross-domain orchestration of services over multiple administrative or technology domains through a Multi-domain Orchestrator (MdO) [5GEx] enables end-to-end network and service elements to mix in multi-vendor heterogeneous technology and resource environments, realizing a full end-to-end service deployment within a reduced time.

A testbed composed of several testbed sites was set up across Europe, in order to experiment with the MdO at large scale and to measure the service deployment time. The experiment is performed in two steps: first, the service provider defines a new service as a chain of multiple VNFs, and publishes it in a service catalogue; the marketplace. Then a customer 'buys' the service from the marketplace, and the MdO deploys the corresponding service chain across the testbed sites. Each MdO can be an entry point for customer service requests (Figure 5 is just a simplified example). If the resources to provision the service are not available in the same administrative domain, the MdO seeks them in its neighbour domain by communicating with its peer MdO. This process can be cascaded through more domains until all the resources have been allocated. The service deployment time is evaluated as the sum of the time the MdO needs for creating the service and publishing it in the marketplace (service creation), plus the time to instantiate the service across the testbed sites upon a customer request (service provisioning). The objective of the experiment is to demonstrate that the MdO is able to reduce the service deployment time from hours to seconds. [5GEx].

FIGURE 5

Multi-domain orchestration for a flexible service deployment.



Reducing service deployment time within the technology domain

Within the domain of the same operator, the service deployment time depends on the availability of physical and software infrastructure and, once the infrastructure is ready, on the performance of the network management and control framework.

First, when the demanded service does not yet exist, a Service Development Kit (SDK) helps reducing development time based on two capabilities: (i) the possibility of reusing VNFs previously validated and stored in a function catalogue in the construction of new services, and (ii) tools such as profiler or emulator, that enable a feedback cycle to the developer of the service, who can test and optimise the service being developed before deployment. Speed of deployment is achieved by providing a service platform that is highly configurable, with an architecture based on micro-services and a message broker that can accept service or function specific managers, which can adapt the default behaviour of the platform [Sonata].

Once the software service is available, one has to ensure that the physical infrastructure for the targeted service coverage is deployed and configured for the desired service. A self-organized autonomic network management framework enables the reduction of the average service creation time through key innovations such as: automated physical, virtual infrastructures, and services deployment; and integrated management and orchestration of SDN/NFV apps for on-demand service creation [Selfnet]. For example, for the use case of provisioning a new antenna in the edge up to its operational stage, measurements have been made to exemplify the achievable service deployment times. The approach being implemented for testing the framework performance is gathering the results for the deployment time for a completely "empty" set of nodes including the deployment of a base station service, supported by an existing OpenStack infrastructure. Time figures to deploy 6 physical machines simultaneously have been gathered in a process that includes all necessary steps, from edge installation to configuring the Evolved Packet Core (EPC), and compared to the figures obtained from a proof of concept prototype. The total time has been reduced from 120 hours to 90 minutes [Selfnet].

Light weight virtualization for a fast VNF deployment

NFV can be used to support highly dynamic scenarios, in which the VNFs are instantiated "on the fly" following the service requests. VNFs tend to become small and highly specialized Micro-VNFs, i.e., elementary and reusable network elements. Complex services can be built through the "chaining" of these Micro-VNFs. Different virtualization approaches can be used to support these micro-VNFs: Tinified VMs and unikernels. Unikernels have very important properties allowing to reduce the service deployment. They offer very good performance in terms of low memory footprint and instantiation time. They have very good isolation and security properties. The recent measurements using ClickOS, a Xen-based unikernel, demonstrate a small footprint (around 5 MB when running) and an instantiation time around 30 milliseconds while processing up to 10Gb/s of traffic. It furthermore does not need a disk to work. In this way, it is possible to make very efficient use of resources, allowing thousands of unikernels to run on a single physical host and offer a fast end-to-end service deployment [Superfluidity].

Network programmability for a reduced service configuration time

The introduction of Software-defined Mobile Network Control (SDMC) in 5G networks extends the concept of network programmability beyond SDN. While SDN decouples network control and forwarding functions, SDMC introduces the separation of logic and agent for any network function in the network, extending the SDN principles to all control-plane, user-plane and management functions typically deployed in mobile networks.

SDMC offers a high level interface to several network functions ranging from radio control to traffic steering. With SDMC, service providers are able to configure the equipment to their needs by simply re-programming the controller using well-defined APIs, enabling a new service within a reduced implementation, test and deployment timescale. The definition of a standard northbound interface simplifies the creation of new network functions, as the low level and vendor-specific characteristics are managed by the SDMC controller southbound interface, with a clear advantage in heterogeneous and dense wireless networks. SDMC can employ tailored algorithms per network slice they are deployed in and can also manage on-the-fly deployment of VNFs close to the users reducing their experienced latency. This feature is desirable for the verticals market, as several network operators can provide their services to verticals by using the SDMC approach.

04 RECOMMENDATIONS

Sections 2 and 3 provided details of technological and architectural innovations researched and developed by 5G PPP projects and covering innovation areas such as:

- Flexible air interface design that serves the three main services classes for 5G
- Integration of different radio access technologies within the 5G transport network
- Flexible architecture that allows quick setup of slices on multi-tenant networks
- Integration of fog computing and mobile edge computing near the end user and within vertical premises.

These innovations are shared with the global 5G research community in order to avoid fragmentation of standards and deployed technologies. This is being achieved by the 5G PPP through international cooperation with other regions such as the Americas, China, Japan, and South Korea. The 5G PPP projects are building pre-standards consensus among their partners and provide relevant contributions to standardization. The 5G Infrastructure Association is a market representation partner of 3GPP that brings into 3GPP a consensus view of market requirements and contributes to the roadmap of 3GPP. Action towards regulation bodies is being undertaken for securing sufficient spectrum for 5G and a stable regulatory framework across all stakeholders. Last but not least, cost effective deployment strategies are being developed, including an upgrade path from 4G, a cost-effective transport network and deployment and operation cost sharing strategies among involved stakeholders.

4.1 STANDARDISATION

Global standardization is an important element of the long term sustainability and the widest possible use of the 5G PPP results and reduces the risk of fragmentation of future deployments. The benefit of standards is that they foster a wide ecosystem and provide increased efficiency by eliminating redundant options. However, standards can limit innovation by preventing use of new technology and raising entry barriers. Therefore, it is important to select the right areas to standardise and create standards with the right properties.

Standardisation has provided profound benefits by the definition of a common air interface and will likely do so in the future. The virtualisation of the network and possible sharing of resources implies the need for standardised interfaces between the virtual functions and the execution platform. This must also cover distributed network approaches where some resources are located at the network edge. Management of future networks is likely more complex and a standardised approach may be beneficial to efficiently manage them. Initiatives such as Multiple Operator Core Networks (MOCN) constitute an enabler for sharing RAN resources. Such initiatives need to be extended covering the entire ICT edge infrastructure.

Standards must be flexible to support and sustain the diversity of business models and deployments of 5G networks. The standards should cover all use case classes; for example it is important that eMBB, URLLC and mMTC are covered in the same standards framework. Standards should be unified and non-fragmented to ensure global and cost-effective mobility of users and equipment.

The 5G PPP programme has influenced the current standards evolution by catalysing the vision what 5G should be about (eMBB, MTC and URLLC) and articulating the overall 5G key performance indicators.

Results to date have been proposed to related standards bodies, most prominently represented by 3GPP, ETSI and ITU. The contributions of results include:

- In 3GPP-RAN specifications of the physical layer of the radio Interface for UE as well as radio interface architecture and protocols, radio resource control and management and the services provided to the upper layers.
- In 3GPP-SA specifications of services and features, definition and evolution of the overall architecture, and addressing security and privacy by design.
- In ETSI contributions to MEC (Mobile Edge Computing), RRS (Reconfigurable Radio System) and TC CYBER (Cyber Security).
- In ITU contributions to SG15 on network technologies for transport.

In the future it should be ensured that use-cases important for Europe are considered and included in the standards, so that the technology required to realise them is developed. Finally, it is important that demand for new services requiring new technology is stimulated so that the equipment is developed and brought to market, fostering a healthy ecosystem.

4.2 SPECTRUM

Early access to the necessary frequency bands is critical for Europe to perform 5G technology tests, trials, pilots and for the early launch of commercial products services. The news release of the Radio Spectrum Policy Group (RSPG) states the following on 5G "pioneer bands"[6] :

- Low bandwidth spectrum (700 MHz) which can enable 5G coverage to all areas, ensuring that everyone benefits;
- Medium bandwidth spectrum (3.4-3.8 GHz) which will bring the necessary capacity for new 5G services in urban areas; and
- High bandwidth spectrum (26 GHz) to give ultra-high capacity for innovative new services, enabling new business models and sectors of the economy to benefit from 5G.

Those RSPG nominated 5G pioneer bands eases the early development of the necessary equipment and facilitates faster preparation of tests and trials.

Sufficiently large radio channels and bandwidths are necessary for supporting the eMBB use case classes at 3.4-3.8 GHz, in addition to the 26 GHz band.

The radio frequency channels needed for 5G are of at least 100 MHz width in the 3 - 4 GHz range rising to 500 MHz in the frequency range between 5 to 33 GHz and as wide as 1000 MHz at the highest mm-wave options.

6 RSPG16-032 "Strategic roadmap towards 5G for Europe - Opinion on spectrum related aspects for next-generation wireless systems (5G)", Brussels, 09 November 2016. See also RSPG Chair News Release on 5G Spectrum, 10 November 2016

4.3
DEPLOYMENT

The key innovations in 5G depicted in this paper represent a revolution for user experience, new services and new business models. The architecture and protocols are designed to adapt to a wide range of deployment scenarios including deep indoor, hot spots, urban areas, rural areas, maritime areas and in an aeronautical context. The 5G concept combines various access technologies, such as cellular, wireless, satellite and wireline, for delivering reliable performance for critical communications and improve area coverage.

The 5G network deployment has to be cost effective in order to materialise. Efficient and progressive deployment strategies, reusing as much as possible 4G infrastructure and exploiting new available spectrum are currently being elaborated in the 5G PPP programme [METIS-II]. For example, in dense urban scenarios, 5G radio base stations should be co-located with 4G base stations and exploit the newly available spectrum, e.g., at 700 MHz and 3.4-3.8 GHz. Furthermore, in order to provide enhanced capacity, small cells could be added in frequency bands below 6 GHz (such as 2.6 GHz and 3.4-3.8 GHz) and in mm-wave bands above 24 GHz. Additionally, complementary use of unlicensed spectrum and the use of nomadic nodes (see section 2.2) are under consideration for increasing the average user throughput.

The design of the fronthaul and the functional split have high impact on the network cost and must be optimised. For example, for a spectrum bandwidth of 200 MHz and beamforming (8x8 antennas) and if the 5G radio unit is in charge of functionalities from physical layer to resource element mapping, the peak throughput for the new radio splitting interface is estimated in some tens of Gbit/s, generating several hundreds of Gbit/s for the aggregate signal in fibre [5GCrosshaul]. Current 100 Gbit/s optical interfaces are too expensive for the fronthaul segment and cost effective 100 Gbit/s direct detection transceivers for Datacom applications cannot provide sufficient link capacity over desired distances. Such interfaces are suitable for peer-to-peer connectivity but are not able to support aggregated traffic in networking scenarios. In view of these constraints and considering realistic number of add/drop nodes in the optical network, two potential solutions are discussed: (i) an optimal functional split that minimizes the Total Cost of Ownership (TCO) [METIS-II] and (ii) novel technologies, such as integrated optical chipsets for fibre dispersion compensation and advanced modulation formats to meet 5G transport requirements at acceptable cost [5GCrossHaul].

The cost minimization strategies alone will not be sufficient for ensuring an economically viable 5G network. Considering the new business models, innovative deployment strategies must be conceived that involve all stakeholders, from telecommunications, content and vertical industries, and which collaborate for sharing the deployment cost and the associated revenue [METIS-II]. The broker business role introduced in Table 1 facilitates resource sharing and makes it dynamic in time and space, exploiting thus networking resources from various access technologies (such as cellular, wireless, satellite and wireline). The trustworthy and secure slicing concept and the multi-domain orchestrator presented in section 3.4 enable the provisioning of an end-to-end service spanning the infrastructure of multiple stakeholders, ensuring reliability for critical communications and improving availability with wide area coverage.

4.4
REGULATION

Regulation must adopt a facilitating harmonised approach for supporting 5G deployment. The promotion of investments requires a stable, consistent and accurate regulatory framework across all stakeholders. Regulation must increase the consistency of administrative conditions to facilitate dense cell deployments, including (i) right-of-way to passive facilities; (ii) supportive municipal site rental charges; (iii) removal of taxation on sites; and (iv) predictable and harmonised electromagnetic field emissions limits.

Regulation rules must promote 5G services and avoid restricting implementation options such as with respect to the slicing concept, end-to-end virtualization and network sharing. Without this flexibility, multi-domain slicing and networking and processing resource sharing cannot be implemented, reducing the economic value for a wide range of services.

Regulation must ensure equivalent and proportionate privacy requirements between operators and online service providers, and remove any roadblocks to development of innovative 5G services in all vertical industries. For low-latency high-reliability 5G services, regulation must clarify the liability issues. Privacy aspects require timely and targeted actions, as stressed by EU Privacy Mandates (e.g. M/530) and the General Data Protection Regulation (GDPR) which requires privacy by design.

Finally the public sector should act as an early adopter and promoter of 5G technologies, for example through public procurement (e.g. for allowing vertical sectors to adopt 5G). Such initiatives help building the business case for the necessary investments.

CONTRIBUTORS LIST

EDITOR

Salah Eddine El Ayoubi
Orange, salaheddine.elayoubi@orange.com

AUTHORS

Jean Sébastien Bedo (Orange)
Salah Eddine El Ayoubi (Orange)
Miltiadis Filippou (Intel)
Anastasius Gavras (Eurescom)
Domenico Giustiniano (Imdea)
Paola Iovanna (Ericsson)
Antonio Manzalini (Telecom Italia)
Olav Queseth (Ericsson)
Theodoros Rokkas (Incites)
Mike Surridge (University of Southampton)
Terje Tjelta (Telenor)

CONTRIBUTORS

George Agapiou (OTE research), **Ricardo Marco Alaez** (University of the West Scotland) , **Jose Alcaraz-Calero** (University West Of Scotland), **Maria Barros Weiss** (Eurescom), **Pascal Bisson** (Thales), **Nicola Blefari Melazzi** (University of Rome Tor Vergata), **José Bonnet** (Altice Labs), **Michael Bredel** (NEC), **Ömer Bulakci** (Huawei), **Carolina Canales** (Ericsson), **Sonia Castro Carrillo** (ATOS), **Fabio Cavaliere** (Ericsson), **Tao Chen** (VTT), **Enrique Chirivella-Perez** (University West Of Scotland), **Nicolas Chuberre** (Thales AleniaSpace), **Alessandro Colazzo** (Azcom), **Luis Miguel Contreras Murillo** (Telefonica), **Giovanna D'Aria** (Telecom Italia), **Thomas Deiss** (Nokia), **Panagiotis Demestichas** (University of Piraeus), **Marco Di Girolamo** (HPE), **Mark Doll** (Nokia), **Manos Dramitinos** (Athens University of Economics and Business), **Simon Fletcher** (RealWireless), **Vassilis Foteinos** (Wings ICT Solutions), **Alexander Geurtz** (SES), **Marco Gramaglia** (UC3M), **Jesus Gutierrez Teran** (IHP Microelectronics), **Frederic Kermel** (Airbus), **Michael Kraemer** (Intel), **Vivek Kulkarni** (Siemens), **Orestis Liakopoulos** (Wings ICT Solutions), **Hakon Lonsethagen** (Telenor), **Diego R. Lopez** (Telefonica), **Jian Luo** (Huawei), **Aarne Mämmelä** (VTT), **Patrick Marsch** (Nokia), **Josep Martrat** (ATOS), **Francesco Mauro** (Telecom Italia), **Ioannis Neokosmidis** (Incites), **Navid Nikaein** (Eurecom), **Miquel Payaró** (CTTC), **Rauno Ruismaki** (Nokia), **Dario Sabella** (Telecom Italia), **Mehrdad Shariat** (Samsung), **Martin Schubert** (Huawei), **Christoph Thuemmler** (Edinburgh Napier University), **Fabio Ubaldi** (Ericsson), **Panagiotis Vlacheas** (Wings ICT Solutions), **Sark Vladica** (IHP Microelectronics), **Thomas Walter** (Docomo), **Qi Wang** (University West Of Scotland), **Yue Wang** (Samsung), **Jean-Philippe Wary** (Orange), **Mick Wilson** (Fujitsu)

5G PPP

5G IA

Supported by the

European Commission

This material has been designed and printed with support from the Euro-5G project and the 5G Infrastructure Association. The Euro-5G Project has received funding by the European Commission's Horizon 2020 Programme under the grant agreement number: 671617.

More information at
www.5g-ppp.eu