Big Data in 5G Distributed Applications

Valentina Nejkovic[1(✉)], Ari Visa[2], Milorad Tosic[1], Nenad Petrovic[1],
Mikko Valkama[2], Mike Koivisto[2], Jukka Talvitie[2], Svetozar Rancic[3],
Daniel Grzonka[4], Jacek Tchorzewski[4], Pierre Kuonen[5],
and Francisco Gortazar[6]

[1] Faculty of Electronic Engineering, University of Nis, Nis, Serbia
valentina@elfak.ni.ac.rs
[2] Faculty of Computing and Electrical Engineering,
Tampere University of Technology, Tampere, Finland
[3] Faculty of Science and Mathematics, University of Nis, Nis, Serbia
[4] Cracow University of Technology, Cracow, Poland
[5] University of Applied Sciences of Western Switzerland, Fribourg, Switzerland
[6] Rey Juan Carlos University, Madrid, Spain

Abstract. Fifth generation mobile networks (5G) will rather supplement than replace current 4G networks by dramatically improving their bandwidth, capacity and reliability. This way, much more demanding use cases that simply are not achievable with today's networks will become reality - from home entertainment, to product manufacturing and healthcare. However, many of them rely on Internet of Things (IoT) devices equipped with low-cost transmitters and sensors that generate enormous amount of data about their environment. Therefore, due to large scale of 5G systems, combined with their inherent complexity and heterogeneity, Big Data and analysis techniques are considered as one of the main enablers of future mobile networks. In this work, we recognize 5G use cases from various application domains and list the basic requirements for their development and realization.

Keywords: 5G · Big Data · Use cases

1 Introduction

The vision of 5G is becoming clearer as we move closer to the end of this decade. The 5G will feature increased network speed, and machines, cars, city infrastructure beside people will be connected. It is expected that 5G networks will have always-on capabilities and to be energy efficient, which require new protocols and access technologies. The 5G network represents highly complex and heterogeneous network that integrates massive amount of sensor nodes and diversity of devices such as macro and small cells with different radio access technologies such as GSM, WCDMA, LTE, and Wi-Fi that coexist with one another. Such network vision is expected to lead to traffic volume of tens of exabytes per month that further demands networks capacity 1000 times higher than now [1, 2]. Such traffic volume is not supported with nowadays cellular networks. Thus, practical deployment of 5G networking systems, in addition to traditional technology drivers, needs some new critical issues to be resolved on different areas

such as: (1) coordination mechanism [3], (2) power consumption [4], (3) networking behavior prediction [5], (4) positioning and location-awareness [6] etc. Some operators already start their deployments and the standards process forward.

5G will be built upon the existing 4G LTE networks with features available as part of the LTE - Advanced standard. Some features will include carrier aggregation that enable using of existing spectrum efficiently with network capacity increase and higher throughput rates. Self-organizing networks will play key role as well as technologies such as coordinated multipoint that will enable operators to simultaneously transmit and process signals from multiple sites. Software-defined networks (SDN) and network functions virtualization (NFV) will be very important for operators in order to scale their networks quickly in migration from 4G to 5G [7]. SDN will play key role for carving virtual sub-networks, which can be used for huge bandwidth applications, which for example include video with requirement in speed of 10 Gb/s as well as lower bandwidth applications, which for example connect different user equipment that are less demanding on the network.

The 5G architecture and deployment will depend upon how the network is used. For example, applications such as streaming video, video conferencing and virtual reality require high speed with growth in the video traffic. In order to achieve this requirement, the network needs a lot of small cell coverage and higher bandwidth spectrum. Further, 5G will be the network for Internet of Things (IoT) with support for a lot of devices [8]. Such IoT network should be efficient in low-bandwidth transmissions with enhanced coverage. Because of the high scale of 5G systems combined with their inherent complexity and heterogeneity, Big Data techniques and analysis will be one of the main enablers of the new 5G critical issues.

Big Data refers to large data sets whose size is growing at enormous speed making it difficult to handle and manage them using the traditional techniques and software tools. It is a step forward from traditional data analysis, considering the following aspects (so-called five "five Vs") [9]: quantity of data (volume), different types of semi-structured and unstructured data (variety), the rate with which data is changing or how often it is created (velocity), the importance of results extracted from data (value), data quality, including trust, credibility and integrity (veracity). Taking into account the prediction that the number of connected devices will increase 10–100 by the time when 5G will be commercially used [10], it can be concluded that Big Data techniques will play an important role, as all the considered usage scenarios are based on extracting knowledge from the enormous amount of heterogeneous data generated by connected devices in order to support the decisioning and other mechanisms in future 5G networks.

In this chapter, we identify use cases and scenarios that could benefit from new capabilities provided by 5G network in synergy with Big Data technologies, list basic requirements for their application development, and consider some future challenges addressing positioning challenges and semantic-based solutions. The researcher community and service providers (business stakeholders) could benefit from this chapter. From one side, it provides an insight of recent trends in 5G research and development, while, from the other side, it discusses how the research outcomes could be used for development of future services to satisfy customer demands.

Section 2 of this chapter gives list 5G use cases requirements, while Sect. 3 gives identified 5g use case with short description of each. Section 4 gives future challenges targeting positioning systems, semantic based approaches, 5G security etc. Section 5 concludes the chapter.

2 5G Use Cases Requirements

The introduction of big data techniques in 5G distributed applications poses a challenge, as these techniques usually require huge computational resources. In general, there is a need for high performance computing infrastructure. This infrastructure would typically be available as a private or public cloud or grid. Cloud or grid resources will allow for consuming, storing and processing huge amounts of data. This data shall be prepared for consumption from the edge network. Depending on the nature of the data needed by end-users, we can envision two kinds of data processing methods: online and offline. Offline methods are easier to handle as the processing can be performed in the cloud or grid. This are supposed to be processes that are not critical. Online processing is used when a response is needed in a given amount of time, and therefore both the time required to give a response and the latency would have high impact on the set of use cases that will be available for 5G.

For online processing, in those cases where common off-the-shelf hardware is available at the edge, general Big Data solutions can be run on top of this commodity hardware, assuming that the constrained resources available are enough.

In general, we identify following requirements needed for 5G use cases:

- Network requirements: Network with 5G capabilities; faster and higher-capacity networks, which can deliver video and other content-rich services; massive connectivity of devices based on different technologies, etc.
- Application requirements: Consistent process mining over Big Data triple store; Network capability measurement module; Reasoning module; Learning and Prediction module (for example, Neural Network); Optimization module; Corresponding domain and application Ontologies; etc.
- Storage requirements: Big Data triple store; Possibility to handles large amounts (a petabyte or more) of data; Distributed redundant data storage; Massively parallel processing; Provides Semantic Big Data processing capabilities; Centrally managed and orchestrated.

Even though the 5G networks are primarily designed for enhanced communication purposes, high-accuracy positioning has been considered as one of the key features in 5G. Moreover, the standardization organization third generation partnership project (3GPP) has already published several technical reports and specifications regarding positioning in future 5G networks [11–14]. However, since the 5G specifications are still under development, detailed descriptions of different positioning approaches and related positioning protocols are yet unavailable. Despite this, in order to facilitate development of various future 5G-enabled use cases, 3GPP has introduced the first set of performance requirements considering different types of position-reliant use cases presented in [13]. For each use case, a specific positioning accuracy, including both

horizontal and vertical directions, has been given according to the type of the area of interest and its 5G service characteristics. For certain use cases, when applicable, also accuracy requirements for velocity estimation and device bearing estimation are provided.

Similar to common 5G guidelines, besides focusing only on maximizing the positioning accuracy, also other important positioning service aspects have been considered in the existing reports given in [13]. One of the key performance indicators is positioning availability, which defines in which percent of the time the positioning method provides estimates with the specified accuracy level. Another important positioning performance indicator is latency, which indicates the elapsed time between triggering the positioning process and finally obtaining the position estimates. Moreover, the latency of the first position estimate at the initialization stage of the positioning process, referred to as the time-to-first-fix, has been separately specified typically with reduced performance requirements compared to the latency in general.

In addition to the above-described 5G positioning requirements, there are various other aspects, which have not yet been appropriately addressed in the reports and specifications, but should be considered according to the needs of users, operators and 3rd parties. Such aspects include, for example, energy consumption, security and privacy, estimation reliability and related confidence levels, and possible regulatory requirements (e.g., positioning during emergency calls) [12–14].

Mobile-network-based positioning approaches can be divided into two fundamental categories, which are network-centric positioning and user-centric positioning. In network-centric positioning, the position estimates are obtained at the network side based on the signals transmitted by the user device. In this approach, all heavy computational load is located at the network side, which reduces the power consumption of the user device, and thus, increases the valuable battery life of the device. Moreover, when the positioning is done at the network side, all positioning related information, such as network BS locations, are already available for the positioning algorithms without introducing additional overhead from signaling the information over the radio interface. The fact that the position information is fundamentally located at the network side is especially useful for achieving the future targets of the 5G networks, as it facilitates numerous 5G-enabled mission-critical use cases where the latency and reliability of the position estimates are in a crucial role. In the user-centric positioning approach, where the user device performs the positioning based on the signals transmitted by the network nodes, the position information is not directly available at the network side. This approach increases user security and privacy, but on the other hand, it requires additional signaling overhead in order to utilize the device positions jointly as part of new 5G-enabled infrastructures such as traffic control and ITS, for instance.

3   5G Use Cases

In this section, we analyze how the current advances in 5G and related concepts can be leveraged in order to provide novel use cases that were not possible before or improve the existing services and solutions.

3.1   5G Coordination Mechanisms

The 5G network indicates the need for coexistence of multiple wireless technologies in the same environment [3, 15, 16]. The problem that raises in such environments is mutual interference among multiple wireless networks, which is consequence of an overlapping in usage of the same set of resources. Typically, such case happens when same radio frequencies are used for multiple communication channels that are based on different radio technologies [3, 15]. Coordination protocols defined by the technology standards traditionally address the problem when networks use same technology. New coordination concepts are needed in the case of co-existing networks based on heterogeneous technologies [16, 17].

We identify the following possible scenario. In a Home Network setting, a typical home can have several rooms each equipped with WiFi enabled HDTV set and a number of streaming audio appliances. At the same time and in the same building, a sensor network is used for home automation including presence detection, temperature and lighting regulation, doorbell indication and security and safety monitoring. Most homes also have at least one microwave oven and a number of Bluetooth Low Energy gadgets. During the typical evening, all of these devices are active and have to be actively coordinated in order to provide satisfactory level of service.

3.2   Power Consumption

A number of recently finished as well as currently on-going 5G related EU projects confirm a diversity of usage and applications of power consumption, efficiency and reliability in WSNs. These projects delivered a number of algorithms and protocols for reducing energy consumption that show the importance of the power consumption. Further, the design of the 5G wireless networks has to consider energy efficiency as very important pillar in order to optimize economic, operational, and environmental concerns [1, 18]. In presence of enormous high traffic volume, data-driven techniques such as intelligent distribution of frequently accessed content over the network nodes and content caching can result in relevant energy consumption reductions and prolong the lifetime of nodes that are low on battery energy.

3.3   Networking Behavior Prediction

Big Data Analytics solutions can predict how the needs in resources use change among places and throughout the time within a complex large-scale system. A 5G network that adopt such solution would have ability to learn from the previous situations and states and intelligently adopt to new demands [5, 19]. Particularly, using appropriate learning techniques the system will enable devices to learn from past observations in their surroundings.

For example, we identify the following use case scenario: (a) In a Smart City, traffic lights and pedestrian crossings (i.e. various presence detectors) are IEEE 802.15.4 technology equipped while community WiFi network is mounted on a number of light posts lining the same street. During rush hours, there is a high demand for WiFi traffic due to a large number of people using personal devices potentially impacting traffic

management system; (b) Mobile users consume images, videos and music, which increase in volume over time. In such a case, network congestion is a consequence of the high dynamics in demands that exceeds the system potential for adaptability.

3.4 Positioning and Location-Awareness in Future 5G Networks

The world is changing rapidly. New services are needed and many of those new services require location-awareness. Autonomous vehicles, transportation, traffic control need this kind of service. If we consider the problem from the point of view of the smart city we notice that there are many new user groups, such as pedestrians, cleaning and maintenance services, management and administration. There are several approaches to help with positioning. One started with Long Range Positioning systems like Decca and LORAN and continued with Global Positioning System (GPS) that is a positioning system based on Medium Earth Orbit satellites. GPS is a part of Global Navigation Satellite System (GNSS) [20]. GNSS also includes, for example European Galileo and Russian GLONASS. However, satellites are not solving the positioning problem totally. In many regions positioning needs help from mobile communication networks that is called assisted GPS (A-GPS). The high latitudes in North and in South and cities with skyscrapers are for examples problematic regions.

In contrast to the earlier and existing mobile generations, where positioning has been only an add-on feature, future 5G radio networks will allow for highly accurate positioning not only for personal navigation purposes but also for unforeseen location-aware services and applications like robotics, intelligent transportation systems (ITSs), and drones, just to name few. While seeking to meet the demanding communication requirements of 5G, e.g., in terms of capacity and data-rates, 5G networks will exploit large bandwidths and massive antenna arrays, which together with even denser base station (BS) deployments create also a convenient environment for 5G-based radio positioning. Hence, it is widely expected that future 5G networks should enable and even improve indoor and outdoor positioning techniques embedded to a radio access network (RAN) [11] as well as the ones that utilizes RAN-external measurements from GNSS or sensors.

3.5 Ultra/High Definition Live Video Streaming in Wireless Networks

In recent years, video streaming, both on-demand and live has become an important part of our everyday lives – from social networks, content delivery platforms to industrial, robotic and experimentation systems. Due to rise of processor power and camera sensor resolution of consumer devices, such as smartphones, the image quality criteria perceived by consumers has dramatically increased. High Definition video is becoming a must for all the use cases where video streaming is involved. Not only that, but also new video formats are emerging, such as stereoscopic 3D, 360-degree video and Ultra High Definition Video which contain even more data that has to be transmitted. Therefore, Internet service providers, mobile carriers and content providers are encountering many issues, as transmission of such content requires significantly larger bandwidth. Additionally, the issues become even more challenging due to device mobility, which can affect the Quality of Service and Quality of Experience, especially

when it comes to live video broadcast in varying network conditions [21]. Here, we identify a potential use case of novel networking paradigms – SDN and VNF, in combination with Big Data technologies. Large amount of network equipment and status data is analyzed. The results of data analysis are semantically annotated and stored into RDF triple store, so semantic reasoning can be performed in order to draw new conclusions which could lead to re-deployment of virtual networking assets, generation of SDN rules, parameter tuning or other optimizations with objective to satisfy user-defined QoS parameters and maintain the quality of high definition live video streaming in varying network conditions, where devices are moving intensively (such as mobile robotic and experimentation systems).

In several publications so far, this topic has been discussed, problems identified and several solutions proposed. However, in most cases, these solutions suffer from low quality, large end-to-end latency in live streaming and frequent freezes in the video playout due to sudden drops of the available bandwidth [22]. In [22], it was shown network-based prioritization introduced by an OpenFlow, SDN-enabled controller can reduce video freezes caused by network congestion. Therefore, the utilization of SDN technologies in this case seems promising. In [23], results confirm that it is now possible to realize a short-range THz wireless communication system for commercial applications where Ultra HD video streaming is needed. However, it is not suitable for use cases like experimentation and mobile robot where long-range wireless communication is of utmost importance.

We can conclude that there are still many open questions in case of ultra/high definition live video streaming using wireless networks, which makes it suitable for future research and application of next generation networking in synergy with Big Data and semantic technologies.

3.6   Multi-party Trust Based on Blockchain for Process Monitoring

IoT and smart objects are key-enabler technologies for monitoring of complex business processes [24], especially in logistics domain and industrial production systems [25]. However, most of these processes involve multiple parties. In absence of central authority, the trust between these parties becomes an important issue [25, 26]. Despite the fact that artifact-driven monitoring enables to effectively keep track of the execution of processes where multiple organizations are involved, it does not fully solve the problem of trust among them. As the devices involved in monitoring process might belong to different organizations, there is still a possibility that one of the parties can misconfigure its devices in order to achieve some own goal in an illegal way, with possibility to disrupt the process execution itself, affecting the final outcome.

Blockchain technology is recognized as a solution for issues related to trust in multi-party process monitoring systems [25–27]. Blockchain provides a shared immutable ledger, which guarantees that the information can be accessed and validated by all the participants of a process both during its execution and after it is completed, which builds the trust among them.

This use case represents a potential area where we can make use of synergy of various novel technologies and paradigms, such as IoT, Big Data and next generation networking together with blockchain.

3.7   Trusted Friend Computing

With the advent of 5G networks, an increasing number of devices will be connected permanently and at high speed to the Internet and to each other. While many applications will benefit from this new connectivity by being able to interact more quickly with data providers, such as the cloud or other resources, there is also a growing need for privacy-conscious data sharing. This is particularly the case in the context of Big Data, which also includes the issue of moving large amounts of private data. One possible approach to this problem is to move calculations close to the data and provide access to both data and local computing resources.

The Trusted Friend Computing (TFC) concept aims to enable a community of users to securely share their IT resources without a central organization collecting and storing information. It is a distributed, resource-centered paradigm where data, computing power, software or the network can be shared reliably and resiliently. This paradigm defines an original IT architecture built around the notion of a community of users (called friends) of a given software application. Instead of using the traditional approach where the IT architecture is middleware-centric to share resources, the TFC approach focuses on the software application used by the community. One of the important advantages of this approach is to avoid heavy executable codes transfers since all friends already possess the calculation modules useful for the community. Inspired by the social network model and using a concept similar to virtual private networks (VPNs), the idea is to allow friends to invite other users of the software to join the community to share their resources. The community is therefore built by individual cooptation and is, by nature, distributed, decentralized and elastic.

To achieve this objective, several major technical challenges must be addressed. We can, among other things, mention:

- Clearly define a security model for sharing IT resources in the context of TFC applications;
- The definition of community management and accounting needs;
- The development of a platform to enable and facilitate the implementation of applications that comply with the TFC model.

Finally, a user community focused on using a specific application must be identified and the application must be enhanced with TFC features. One of these communities is that of physicians involved in the diagnosis of genetic diseases and using the Gen-searchNGS tool [28]. This Java software analyzes next-generation sequencing data (NGS) to detect changes in DNA sequences for the diagnosis of genetic diseases [29, 30]. In order to be able to easily integrate into any Java software, including Gen-searchNGS, TFC capabilities the POP-Java tool [31, 32] was used. This tool has been improved to support the different functionalities required for TFC-compatible applications [33].

TFC's security model is based on the notion of a "confidence link" as presented in [34]. A confidence link is a two-way channel that allows two friends to communicate

safely at any time. The confidence link also authenticates users with a security certificate, ensuring the identification of communicating partners. Each member of a network can extend the network by creating additional confidence links with other friends, thus adding new members to the network. All friends as well as all confidence links form a connected graph where the nodes are friends and the arcs are the confidence links. We call such a graph a "community of trusted friends" or more simply a "community". None of the friends in the community have a global view of the infrastructure. Each friend only knows his direct friends, i.e. the users with whom he has established a confidence link.

Applications can publish resources on a network of friends or search and access the resources of other network members. When publishing a resource, specific access rights can be given to limit access to the resource, for example by differentiating between direct and indirect friends in a network.

The model also includes the ability to record the use of each member's resources, which allows the use of resources to be billed based on their utilization rate, thereby encouraging members to share their resources.

Today, and certainly even more so tomorrow, the use of mobile networks for professional applications will be a reality. These applications are less and less confined to work desktops but are now used outside the enterprise environment for efficiency and ease of use. A concept such as TFC can truly benefit from a high-performance mobile communications network such as 5G networks to provide professional communities with secure access, anytime, anywhere, to vast computing and data resources.

3.8   Virtual and Augmented Reality Applications

Virtual (VR) and augmented reality (AR) applications are not exceptions when it comes to potential use cases where utilization of 5G networks could be highly beneficial. The arrival of next-generation of mobile network will unlock the full potential of VR and AR technology, which is still limited by current network characteristics. The complex graphically-rich scenes and sophisticated input mechanisms that are used to create the VR and AR experiences require a large amount of data that has to be processed [35]. Lag, stutter, and stalls are unacceptable for user experience and comfort [36]. This is not a huge problem for local applications, but is quite challenging when done remotely, if the user is on the move and not using the fixed network connection [35]. In this case, the quality of VR and AR experience is heavily dependent on three network components: high capacity, low latency and uniform experience. This way, many novel services and applications that involve the usage of augmented and virtual reality would see lights of the day, such as immersive movies, video games, live shows, concerts, sport events, immersive education platforms, immersive social interactions, immersive professional project collaboration and many others [35–37]. To sum up, these services would affect the way that people play, learn and communicate [36].

4   Current Solutions and Future Challenges

4.1   5G Positioning

Possible Error Sources State of the Art

When considering the 5G positioning aspect, ultra-dense BS deployments, increased transmission bandwidths, and large antenna arrays enable efficient utilization of both ranging-based (e.g., time-of-arrival (ToA) and time-difference-of-arrival (TDoA)), and angle-based (e.g., direction-of-arrival (DoA)) positioning measurements. However, in order to exploit these types of measurements for positioning, specific prior knowledge about the network and user device is often assumed available. In case of temporal measurements, the clocks of the user device and network nodes are often assumed to be synchronized. More specifically, with ToA measurements, all clocks in the network, including the user device and the BSs, are typically assumed to be synchronized among each other, whereas with TDoA measurements, only the BS clocks are assumed to be synchronized. Nonetheless, clock synchronization errors can result in large inaccuracies in ranging measurements, and thus, has to be carefully considered in a practical positioning system implementation. Besides the aforementioned clock errors, the ranging measurements as well as angle-based measurements can be deteriorated by the errors related to BSs' locations. In addition to the inaccurate BSs' location information, uncertainties in the orientation of the BS antennas and/or the user device antennas may cause significant error to the positioning results when utilizing angle-based measurements like DoA measurements for positioning.

Whereas the BS position and antenna orientation error can be typically considered time-invariant, the clock errors are often time-variant with certain time-drifting behavior. However, it is appropriate to assume that the time-variant behavior of the BS clocks can be sufficiently small, and thus, there can be only a possible constant clock offset between the BS clocks. Nonetheless, any unknown (or uncertain) system parameter, such as clock offset, BS positions and antenna orientation, can be estimated using classical simultaneous localization and mapping (SLAM) approaches, where the user device position and the unknown system parameters are estimated simultaneously while the user device is moving within the network coverage area.

Since the 5G specifications are still under development and 5G networks are only beginning to emerge to the market, the state-of-the-art 5G positioning studies rely on high computational load computer simulations using realistic radio wave propagation models with extensive 3D ray tracing algorithms. In [38], a network-centric positioning approach was studied by considering asynchronous clocks in the user device and in the network BSs. It was shown that regardless of the clock errors, sub-meter positioning accuracy was achieved by using the ToA and DoA measurements. Moreover, while the user device was moving in the network, the network BSs were synchronized similar to the well-known SLAM principle. This type of approach was later used in [39] for investigating location-aware communications, including applications for proactive radio resource management and location-based geometric beamforming.

A user-centric positioning approach based on signals from a single BS was studied in [40]. In this case, by utilizing only a single BS for the positioning, requirements for

the clock synchronization can be considerably alleviated. Based on the ToA and angle-of-arrival (AoA) measurements, the user device position was estimated with sub-meter accuracy and the antenna orientation of the user device with sub-degree accuracy. Moreover, the developed estimation algorithm was also designed to exploit reflected (or scattered) radio wave components, and therefore, it was able to provide position estimates also for the reflection locations. This type of utilization of non-line-of-sight radio paths introduces various new communications aspects from advanced beam-forming techniques to environment-aware interference management. The user-centric positioning approach was also studied in [41] for a high-speed train scenario utilizing 5G-specified downlink synchronization signal blocks for positioning purposes. Again, despite of the challenging BS geometry of the train scenario, sub-meter positioning accuracy was achieved by jointly using the ToA and AoA measurements.

The network-centric positioning with uncertain BS antenna orientations was studied in [42], where the positioning was based on type of signals used in conventional beam training procedures. By using beam-wise received signal power measurements the user device position and the unknown BS antenna orientations were jointly estimated achieving a sub-meter positioning error and a sub-degree antenna orientation error.

Semantic Analysis of Network Topology and Sensor Data for High-Precision Localization in 5G Networks Challenges

High-accuracy positioning has been considered as one of the key features of future generation network and still an open question in many areas, such as robotics, drone-based experimentation and exploration, autonomous vehicles and intelligent transportation systems.

This task becomes quite challenging in these cases, especially when it comes to indoor localization [43] and outdoor localization of fast-moving aerial devices [44] in varying network conditions (drones).

Mobile-network-based positioning can be divided in two categories: user-centric and network-centric. There are various positioning methods which perform with different values of accuracy, latency and time-to-fix in certain conditions.

It is identified that current research in 5G localization is going towards cooperation [45, 46]. Therefore, the semantic coordination of both user and network operator devices could be used for determining the precise location, taking into account two factors:

1. Network topology: how the devices are arranged in space within the network, such as distance, frequency band at which the device is operating etc. The information about network topology can be semantically annotated leveraging some domain-specific language as a representation.
2. Service utilization and sensor data: a large amount of service utilization and sensor data is collected from both the customer and network operator devices (such as monitoring and status). It can be analyzed leveraging various data analysis technique. Furthermore, the data can be semantically annotated according to the results obtained as output of data analysis techniques.

For this purpose, we define domain-specific ontologies and rules which are used to perform semantic reasoning about the precise location leveraging the semantic annotations about both the network topology and service utilization/sensor data, taking into account the user-defined performance metrics and QoS parameters, such as accuracy, latency and time-to-fix (Fig. 1).
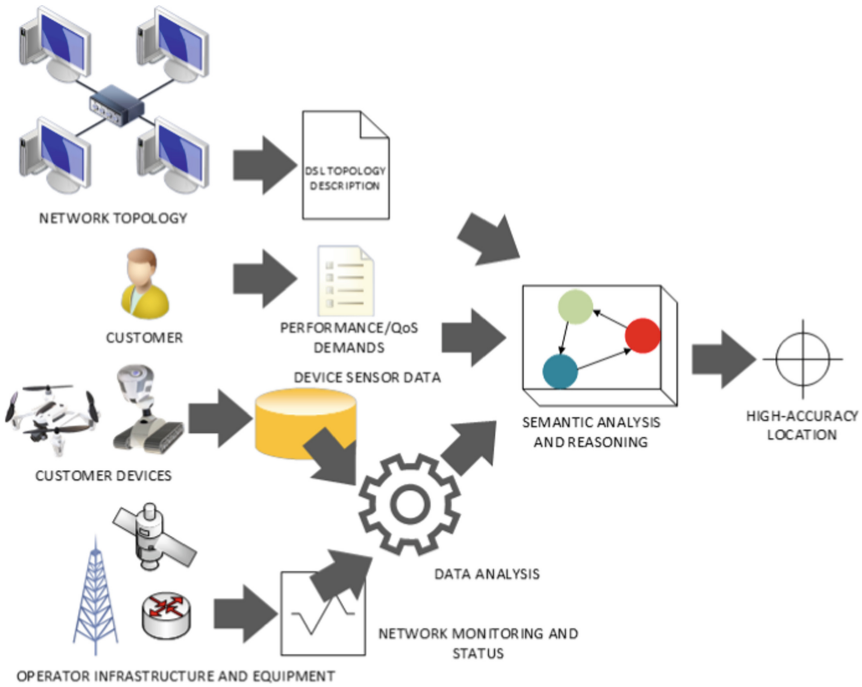

Fig. 1. Semantic analysis of network topology and sensor data for high-precision localization in 5G networks.

4.2   Infrastructure Design of Semantic Driven Big Data in 5G Networking

Semantic Driven Big Data State of the Art
Problems of 5G coordination, power consumption and network behavior prediction feature smart adoption of high-volume data processing results by the system that we propose to address using semantics. In particular, core of the proposed infrastructure is a server, centralized or distributed, that collects relevant knowledge in the given environment and uses the knowledge to make necessary informative decisions, for example about network coordination, network sensors power consumption etc. The server collects networking data and interprets data semantically. For the knowledge representation, the server uses ontology framework approach. The first version of the framework has been previously successfully applied in the case of coordination of technologies that operate in the same unlicensed frequency band [17, 47]. The

coordination and spectrum sensing is modelled as an interactive process, where system nodes communicate and share knowledge about relevant spectrum conditions. Semantic channels are established within the system for the interaction between participating communication devices. The ontology framework could be extended for different cases such as solution presented in [48] that could give further directions for management of semantic Big Data for intelligence.

System that incorporates sensors and 5G user equipment acquire large collection of data. Collecting, storing, analyzing and retrieving data from industrial sensors or other machinery connected to the Internet of Things has become of increasing importance, as a growing number of organizations is looking to take advantage of available data.

One possible semantic framework approach has been successfully proven in many cases. In [17], the case of semantic LTE-U coordination is presented. The coordination and spectrum sensing is modelled as an interactive process, where system nodes communicate and share knowledge about relevant spectrum conditions. Ontologies are used for knowledge representation as bases for automatic reasoning about optimal channel allocations and for coordination. Moreover, in [49] the semantic technology was used for the implementation of network intelligence on top of the FIESTA-IoT platform by using reasoning for the network state estimation in order to perform the spectrum coordination. On the other side, in [50, 51], a semantic-driven approach for unmanned vehicle mission coordination in robotic experimentation testbeds is presented. In this case, the ontologies are used to represent the knowledge about device capabilities, constraints, domain expert knowledge and both the design-time (mission code) and run-time (sensor data) aspects that are taken into account during the generation of the coordinated device missions. Furthermore, in the paper [52] is presented the novel semantic-based approach and algorithm for automatic code generation with huge potential with its extension to 5G applications.

Semantic Driven 5G System Architecture Challenges
The challenge is to exploit semantic technologies at the backend as a flexible foundation for advanced frontend data processing tasks. Possible system architecture consists of five modules given in Fig. 2 and described in more details in the following.

(1) Data Acquisition Module (DAM): Redis is an in-memory database with option of persistence on disk, so it represents a tradeoff where very high write and read speed is achieved at the price of the limitation of data sets that can't be larger than memory [53–55]. We assume data sources at the order of million data series with about million measurements annually for few tens of years with several bytes of a data item size. Hence, size of the total data load could be estimated at the order of a petabyte (PB). The row data has low information density, and as such it is very susceptible for compression. Hence, it can be expected that a 100 times compression rate can be achieved easily (e.g. simple run-length encoding). Big Data technologies can be used for ultra-large scale data sets processing. Distributed storage and processing frameworks are used for that, such as the open source Apache Hadoop Framework [56]. Apache Hadoop enables distributed data processing across clusters of computers. Popular MapReduce distributed data-processing model, Hadoop Distributed File System (HDFS), and distributed table store HBase [57] are Hadoop components.

(2) The Pre-Processing Module (PPM) identifies instability intervals that are semantically annotated, stored, and retrieved later on during search by end user. Anomaly detection is an important problem that has been under extensive research in diverse application domains. We distinguish two basic types of approaches with respect to domain specification as well as online or offline processing. These approaches are not mutually exclusive but, in opposite, they can be used together to achieve a synergy effect. Results obtained in such a way can be used for instant reaction but also for longer term planning activities. They can provide users a valuable information, which can be used proactively, further improve system
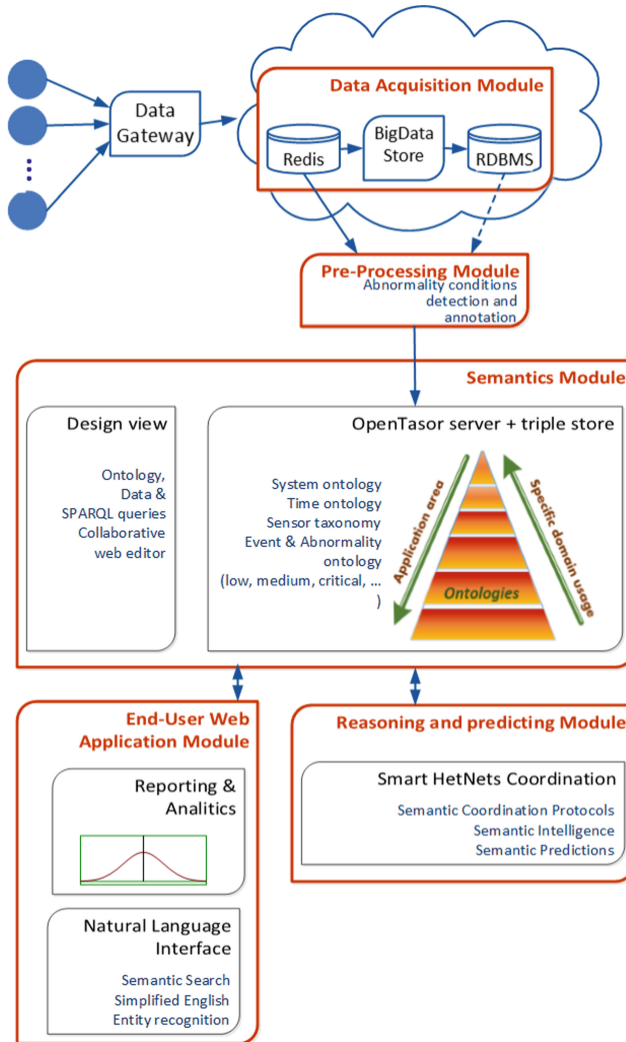

Fig. 2. System architecture for Semantic Driven Big Data in 5G networking.

efficiency and give competitive advantages. In most cases a refining of input data is needed as a first step. Anomaly detection in time series data obtained by sensors [58] is a very demanding task but really important in the same time.

A simple low-complexity algorithm for instability intervals detection is envisioned. The process complexity shall be encapsulated into a separate instability parameters construction module that would extract the parameters from data series in an independent batch-processing manner. Semantic description of the instability interval may be more or less complex depending on the end user application requirements. However, it contains pointer to the corresponding row data record that are stored separately such that semantic search may retrieve row data also. We estimate no more than 250 potential instability intervals within one series annually with average 1% of sensors detecting the instability at one measurement time instant, resulting in $250 \times 0.01 \times 4M = 10M$ instability intervals annually. If we assume semantic annotation of the intervals of 10 triplets per interval, it totals to 100M triplets. This data size is proven to be practically successfully implemented on a off-the-shelf single server hardware with 64 GB of RAM. Note that the system is easily scalable by simple multiplication of the servers assuming federated queries are implemented.

(3) Semantics Module (SM) is based on a platform for scalable linked data semantic datasets management. The platform is envisioned to feature advanced Web based collaborative ontology editor and to be flexible with respect to the used triplestore. By default, we assume a triplestore that is based on Jena [59] as one of the most proven Open Source semantic technologies on the market. Semantic data is represented in standard RDF/OWL formats [60] and manipulated by semantic queries written in the standard SPARQL query language [61]. In this way, the technology would allow different reasoners to be adopted. The expected data processing and storage efficiency is based on the effective use of semantic descriptions of physical characteristics of sensors, their organization and placement, price, type of measurement units, etc. For the purpose a number of standard ontologies may be loaded into the system and used, such as time ontology, measurements ontology, etc. For application specific purposes an online collaborative ontology editor will be used to allow end-user to adjust existing ontologies and develop new ones.

When we have sensor data, attention on the ontologies for sensor data and metadata should be put. The most corresponding is the Semantic Sensor Network (SSN) ontology. Systems that adopt the SSN ontology are built on an RDF database - triple store. Big volume of sensor data collected are challenging to triple stores, because the evaluation of SPARQL queries becomes expensive. Triple stores are not optimized to evaluate time series interval queries. Emrooz is good solution for such case. Emrooz is open source and scalable database capable of consuming SSN observations represented in RDF and evaluating SPARQL queries for SSN observations [62]. Emrooz can be implemented on Apache Cassandra and Sesame [62, 63].

(4) End-User Web Application Module (WAM) is envisioned to be implemented as an advanced Ajax front-end that communicates with back-end using RESTful service API and SPARQL queries. It should consist of the following sub-modules:

(a) NLP Sub-Module: user enters search query in simplified English, such that low cognitive load for end user is required while in the same time certain domain language specifics are exploited in order to lower complexity of the language use and processing. The input is then processed and converted to SPARQL query for semantic search over the semantically annotated data. Data preprocessing algorithms can then be used to retrieve intervals of unstable states, with recorded date and time of the start and end of the data sequence. Every time series is semantically annotated also: name of the physical property, type of the measurement unit, name of the subsystem (and/or machine and/or location …), attached sensor, etc. Semantic descriptions of data are automatically generated during preprocessing and are later used for making autocomplete recommendations to the user, to help him easily search the time series descriptions. Then, key segments are identified related to the given search query, where each of the segments can describe date, time, abnormality in measured physical unit (on some sensor with some features), and all/some unstable data overlapping with an extracted instability interval [64]. In this way, we are able to make more effective and user friendly queries.

(b) Reporting & Analytics Sub-Module: Results of the data analysis are visually presented by means of visually appealing charts, graphs, tables, etc. Semantic filtering is applied for powerful faceted search and browsing through the search results. Also, data analysts are able to reconfigure online data presentation into simple web applications that could be instantly used by the other team members.

Semantic similarity between the input query and annotations of the instability intervals is used as the indicator of coincidence between data sets and the search query. Multilingual support can be provided by existing features defined in the standard RDF/XML language support. As a consequence of the NLP module that is based on representing concepts by the supporting ontologies, multilingualism is supported naturally. Though, some additional effort would be required depending on the type and number of additional languages.

Additional features: Different characterizations and taxonomies of the instability intervals are possible including for example classification of the abnormal events as low, middle, high or critical. We can also specify the intensity of the abnormality as a percentage of the deviation of maximal measured value in the instability interval from the target value. For example, we may define abnormality as "deviation between 7% and 20%". Estimation of conditional probability of overlapping of two or more instability intervals will be based on simple analytics algorithms such as counting ("x of y" or "x/y", 6 of 9 or 7/10, in searching results indicate overlapping of the instability intervals). Advanced user-friendly simplified English based search for causality chains for identification of the root cause is possible. Similarly, a set of intervals rooted by a specified interval can be determined as well.

(5) Reasoning and Predicting Module (RPM) is envisioned to be implemented as an advanced neural network. Neural networks can be adopted for learning component of the RPM. Neural networks have been proven effective in interference detection and classification within wireless networks [65, 66] as well as in time-series prediction tasks [67] that are crucial for coordination.

4.3   5G Security Challenges

Security solutions for 5G can be divided into five groups: Software Defined Network (SDN), Network Function Virtualization (NFV), Mobile Cloud (MC), communication channels and privacy policies [68]. Primary focuses with target technologies described in [68] are: security of centralized control points (SDN, NFV); flow rules verification in SDN switches (SDN); control access to SDN and core network elements (SDN, NFV, MC); isolation for VNFs and virtual slices (NFV); security of control channels (SDN and channels themselves); user identity verification for roaming and clouds services (privacy policies); security of users identity and location (privacy policies); encryption and anti-malware technologies (privacy policies); security of data, storage systems and web services in clouds (MC); service-based access control security for clouds (MC).

Each of this target and security technologies are deeply investigated in [68].

Security of 5G will be a big challenge because it will connect branches of critical infrastructures. However, to make 5G a safe technology, security solutions will consider not only this integrated critical infrastructure but also society as a whole, [68]. The basic challenges mentioned in [68] and [69] are:

- High network traffic - a huge number of IoT devices.
- Security of radio communications.
- Cryptographic integrity of user data plane.
- Roaming Security - updating security parameters between operators of networks.
- Denial of Service and Distributed Denial of Service attacks on infrastructure and end devices.
- Coordination of distributed control systems (like Non-Access Stratum layers of 3GPP protocols).
- Eavesdropping. This attack may lead to intercepting messages by an attacked receiver and is very hard to detect.
- Jamming. This attack may lead to disrupting a communication between legitimate users or block access to radio resources. Very often is realized via an infected receiver.
- Man in The Middle attack. Attacker takes control over communication between legitimate users.
- Basic requirements like authentication, authorization, availability or data confidentiality. Some of the current technologies fulfilling these requirements may be not effective enough in 5G context.

4.4 5G Simulations

A very popular method of finding problems, predicting behavior and developing improvements in system is analyzing the simulation of this system. In 5G cellular communication new simulation systems have to be developed because of a huge number of new services, applications requirements, and performance indicators, [70]. There are three basic types of simulations: link-level, system-level, and network-level. Authors in [70] describe the following challenges connected to all of these three types of simulation:

- Variety of application, technologies, environments and performance indicators.
- Complexity of simulation and simulators. It is caused by growing memory demands and time of simulation which is a result of huge MIMO and complexity of channels.
- Integration of all these three types of simulations. Integration of link-level and system-level simulation may be useful in the evaluation of nonlinear operations (like NOMA) in complex environments. Integration of system-level and network-level simulation is useful in the evaluation of the end to-end performance of all network.

Other challenges, like reusability, scalability, flexibility, multiple levels of abstraction or parallel processing are deeply investigated in [70].

4.5 Radio-Access Research and Propagation Issues

One of the expectations for 5G is to ensure prospective radio-access technologies which will be integrated and will allow creating a long-term networked society [71]. Seven main challenges in this field mentioned in [71] are multi-hop communication, device-to-device communication, cooperative devices, ultra-reliable communication, massive machine-type communication, inter-vehicular/vehicular-to-road communication, and ultra-dense deployments.

The principals of propagation of centimeter waves are very similar to millimeter waves but have different characteristic [72]. The most important differences in those characteristics are free space path loss, diffraction, reflection and scattering, material penetration. These problems can be solved by deploying a Multi Input Single Output System described in [72]. OFDM can be used as a base for developing a new system of encoding digital data on multiple carrier frequencies [72]. OFDM will allow to avoid multipath effect, gain spectral efficiency, and simplify equalization (in comparison with Single-Carrier Systems) [72].

4.6 Millimeter Waves and 5G Standardization

Millimeter waves have a bigger spectrum in cellular frequency bands then centimeters waves [71]. This provides new radio-design opportunities and challenges, like [71]:

- Very high capacity and data rates.
- "Short wavelengths necessitating large array antenna solutions to maintain useful link budgets".
- Antenna sizes will be smaller (design challenges).

Standardization of 5G is a still on-going process. However, some decisions have been already made. The World Radio Communication Conference promoted bands lower than 6 GHz or between 24–84 GHz in 2015 [71]. A Third Generation Partnership Project (3GPP) completed first 5G specification in 2018. Below 6 GHz bandwidth requirements regarding cellular network did not change because of similarity of propagation conditions in new and existing bands [71]. More technical information about higher bands and accepted by 3GPP standards can be found in [71] and [73].

4.7   5G Modulation Schemes

Extreme data rates, a huge number of IoT devices, high-speed high resolution streaming videos - this are only examples what 5G will be used for. The main challenge is to support very fast entry to the network, even for transferring trivial data, [74]. To achieve this goal proper modulation scheme have to be chosen.

Orthogonal Frequency Division Multiplexing is a classic modulation scheme based on dividing available bandwidth into several parallel sub-channels. Each of these sub-channels (called also sub-carriers) can transmit independent data. Multiplexing in time and frequency is possible. However, authors in [74] proposed three modulations schemes, with better Peak to Average Power Ratio and better spectrum efficiency. These modulations schemes are, [74]:

- Filter Bank Multi-carrier (FBMC) - each sub-carrier is filtered independently. Cycle prefix is not used. Offset-QAM is used for orthogonality.
- Generalized Frequency Division Multiplexing (GFDM) adaptable multiple carrier transmission methods. Each sub-carrier is filtered independently. There is no orthogonality. Available spectrum is spread into segments.
- Filtered Orthogonal Frequency Division Multiplexing (F-OFDM) - an extension of classic OFDM. Bandwidth is divided into sub-bands depending on the application. Each sub-band provide proper service. The spectrum is accommodating a range of services which optimize its usage.

Comparison of all these modulation schemes, results and conclusions can be found in [63].

4.8   Machine Learning in Software Defined Networks

The 5G technology entails a significant increase in the amount of processed data. The continuous collection and analysis of such data leads to a Big Data problems that are caused by the volume, variety and velocity properties [75].

However, a key aspect of the operation of each network is its management and control. Recently, most of a network functions (e.g. routing, switching, firewalling, conversion of protocols etc.) were realized by dedicated hardware. The complexity of a network infrastructure increases a number of challenges in organizing, managing and optimizing network operations. The popular idea for solving these problems is Software Defined Networking (SDN) paradigm. SDN allows to migrate many of network functions from the devices to the software-defined networking controllers. The SDN controller manages flow control, analyses network traffic and routes packets according

to forwarding policies. Consequently, SDN controller serves as a sort of operating system for the network.

Taking into account aforementioned challenges and problem of processing large data sets, there is a need for developing efficient and much more complex management methods. Such management methods require making decisions in the real time. There are a lot of known data processing methods. However, many of them cannot be directly applied for effective processing and management of large data sets in modern environments, such as 5G networks. Modern solutions require complex decision making techniques that analyze historical, temporal and frequency network data [76].

One of the possible solutions could be the application of Machine Learning (ML) methods, which are successfully used in the processing of Big Data [77–80]. The capabilities of SDN (e.g. centralized control, global view of the network, software-based network analysis, and dynamic forwarding policies) may fit well to the application of Machine Learning techniques [81]. These possibilities are included in the FCAPS (Fault, Configuration, Accounting, Performance, Security) management ISO standard [82]. In each of the following areas of application one can find intelligent methods [78]:

In the fault management area, ML methods allow not only detection, but also solving the causes of failures in networks. Automation dealing with failures will allow for minimization of downtime and human intervention and, as a result, minimization of losses.

Machine Learning can play important role also in configuration management. Networks such as 5G are characterized by frequent topological changes. This requires modifications in the configuration, which can be prone to errors and difficult to optimize. Considering the multiplicity of configuration parameters, analyses of ML can help to automate this process, e.g. by dynamic resources allocation or services configuration. Appropriate methods can also allow verification of the used configuration and its possible withdrawal and rollback.

Accounting management is tightly connected with monitoring of network resources and pricing plans. ML methods can help identify fraud and dishonest activities of network users. It is also possible to analyze the use of resources and create new service packages. Smart solutions can also significantly improve the QoS level.

An important area of management is performance management. Guaranteeing adequate level of performance is a key factor for efficient network. The use of ML methods can result in traffic load prediction, and, in result, proactive and adaptive network performance management.

Security Management has become crucial issue in networks. Modern security approaches consist of tools for identifying threats and vulnerabilities. The use of ML methods can help in detection of anomalies finding and abuses verification in the network. However, this approach has a high risk of blocking the correct network traffic (high false positive rate). Identifying the nature of the cyber-attack is crucial to choosing appropriate remedies allowed returning to the proper functioning of the network.

The aforementioned opportunities show a wide field for applying ML methods in Software Defined Networking paradigm. Machine Learning can play the major role in autonomous network management for 5G networks. Some of the available solutions, such as IBM's MAPE-K or CogNet, are successfully supported by Machine Learning methods [76, 78].

5   Conclusion

Use case opportunities will increase enormously with 5G networks deployment. Not only that the existing applications and solutions will be enhanced, but many novel use cases and scenarios will become feasible. The potential for further 5G use cases in future services and applications is huge in industries and national priorities including domains from entertainment and telecommunication services, to healthcare, smart cities, remote industrial machine operation, virtual sports attendance and many others.

However, the future scenarios will place much more diverse requirements on the system that need to be explored and analyzed. It is identified that the main enablers of future 5G networks are Internet of Things (IoT), Big Data technologies, together with novel networking paradigms – Software-Defined Networking (SDN) and Network Functions Virtualization (NFV). New architectures will rely on large number of connected smart devices generating enormous amount of data each moment. The generated data needs to be analyzed in order to make the right decision as soon as possible, almost in real time. On the other side, the increased flexibility of network infrastructure management and fine-grained control is also required, which is enabled by NFV and SDN. Furthermore, there is a need for evolution of the current architectures by adding the additional network intelligence layer that would enable more complex scenarios, such as device coordination. The possible approaches for embedding the network intelligence are either using the semantic technology or machine learning techniques.

The future may seem far ahead but the phase for defining the requirements is now. Any new technology or system that we design for 5G needs to be deployed and evaluated, and it is expected to last at least until the end of the next decade.

References

1. Buzzi, S., et al.: A survey of energy-efficient techniques for 5G networks and challenges ahead. IEEE J. Sel. Areas Commun. **34**(4), 697–709 (2016)
2. The 1000x data challenge, Qualcomm, Technical report. http://www.qualcomm.com/1000x. Accessed 29 Jan 2019
3. Yang, L., Zhang, W.: Introduction. In: Yang, L., Zhang, W. (eds.) Interference Coordination for 5G Cellular Networks. SpringerBriefs in Electrical and Computer Engineering, pp. 1–11. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24723-6_1
4. Salem, F.E., et al.: A energy consumption optimization in 5G networks using multilevel beamforming and large scale antenna systems. In: Proceedings of 2016 IEEE Wireless Communications and Networking Conference, pp. 1–6 (2016). https://doi.org/10.1109/wcnc.2016.7564904
5. López, B., et al.: An approach to data analysis in 5G networks. Entropy **19**(74), 1–23 (2017). https://doi.org/10.3390/e19020074
6. Ejaz, W., et al.: Internet of Things (IoT) in 5G wireless communications. IEEE Access **4**, 10310–10314 (2016). https://doi.org/10.1109/ACCESS.2016.2646120
7. Pérez, M., et al.: Self-organizing capabilities in 5G networks: NFV & SDN coordination in a complex use case. In: Proceedings of EuCNC 2018, pp. 1–5 (2018)
8. Khan, M., et al.: Location awareness in 5G networks using RSS measurements for public safety applications. IEEE Access **5**, 21753–21762 (2017)

9. Mohammed, A.F., Humbe, V.T., Chowhan, S.S.: A review of big data environment and its related technologies. In: Proceedings of 2016 International Conference on Information Communication and Embedded Systems (ICICES), pp. 1–5 (2016)
10. Hu, F.: Opportunities in 5G Networks: A Research and Development Perspective. CRC Press, Boca Raton (2016)
11. 3GPP, TS 38.455 (V15.1.0): NG-RAN; NR Positioning Protocol A (NRPPa) (2018)
12. 3GPP, TR 38.913 (v14.3.0): Study on scenarios and requirements for next generation access technologies (2017)
13. 3GPP, TR 22.872 (v16.1.0): Study on positioning use cases; Stage 1 (2018)
14. 3GPP, TR 38.305 (v15.1.0): Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN (2018)
15. Soret, B., et al.: Interference coordination for 5G new radio. IEEE Wirel. Commun. **25**(3), 131–137 (2018)
16. Milosevic, N., Dimitrijevic, B., Drajic, D., Nikolic, Z., Tosic, M.: LTE and WiFi co-existence in 5 GHz unlicensed band. Facta Universitatis, Electron. Energ. **30**(3), 363–373 (2017)
17. Tosic, M., et al.: Semantic coordination protocol for LTE and Wi-Fi coexistence. In: Proceedings of European Conference on Networks and Communications, EUCNC 2016, Athens, Greece, pp. 27–30 (2016)
18. Rizvi, S., et al.: An investigation of energy efficiency in 5G wireless networks. In: Proceedings of 2017 International Conference on Circuits, System and Simulation (ICCSS), pp. 142–145 (2017)
19. Abdalla, S., Ariffin, S.: The enhanced user history-based prediction in 5G. In: Proceedings of the Fourth International Conference on Advances in Computing, Communication and Information Technology - CCIT, pp. 1–5 (2016)
20. What is GNSS? https://www.gsa.europa.eu/european-gnss/what-gnss. Accessed 29 Jan 2019
21. Kacianka, S., Hellwagner, H.: Adaptive video streaming for UAV networks. In: Proceedings of the 7th ACM International Workshop on Mobile Video - MoVid 2015, pp. 25–30 (2015)
22. Petrangeli, S., et al.: Live streaming of 4K ultra-high definition video over the Internet. In: Proceedings of the 7th International Conference on Multimedia Systems Article No. 27, pp. 1–4 (2016). https://doi.org/10.1145/2910017.2910627
23. Nallappan, K., et al.: Live streaming of uncompressed 4K video using Terahertz wireless links. In: Proceedings of 2018 IEEE International Conference on Communications (ICC), pp. 1–7 (2018). https://doi.org/10.1109/icc.2018.8422216
24. Meroni, G., Di Ciccio, C., Mendling, J.: Artifact-driven process monitoring: dynamically binding real-world objects to running processes. In: Proceedings of the Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering, CAiSE 2017, Essen, Germany, pp. 1–9 (2017)
25. Meroni, G., Plebani, P.: Combining artifact-driven monitoring with blockchain: analysis and solutions. In: Matulevičius, R., Dijkman, R. (eds.) CAiSE 2018. LNBIP, vol. 316, pp. 103–114. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92898-2_8
26. Baresi, L., et al.: mArtifact: an artifact-driven process monitoring platform. In: Proceedings of BPM Demo Track Co-Located with 15th International Conference on Business Process Modeling, BPM 2017, Barcelona, Spain, pp. 1–5 (2017)
27. Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., Mendling, J.: Untrusted business process monitoring and execution using blockchain. In: La Rosa, M., Loos, P., Pastor, O. (eds.) BPM 2016. LNCS, vol. 9850, pp. 329–347. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45348-4_19
28. Wolf, B., et al.: DNAseq workflow in a diagnostic context, and an example of a user friendly implementation. J. Biomed. Biotechnol. **2015**(3), 1–11 (2015)

29. Wolf, B., et al.: Safe variant annotation sharing across laboratories. In: Variant Detection 2017, Spain (2017)
30. Böck, J., et al.: Single CpG hypermethylation, allele methylation errors, and decreased expression of multiple tumor suppressor genes in normal body cells of mutation-negative early-onset and high-risk breast cancer patients. Int. J. Cancer **143**, 1416–1425 (2018)
31. POP-Java: Parallel Programming with Java. https://icosys.ch/pop-java. Accessed 29 Jan 2019
32. Wolf, B., Kuonen, P., Dandekar, T.: POP-Java: Parallélisme et distribution orienté objet. In: Proceedings of the Conférence d'informatique en Parallélisme, Architecture et Système, Compas2014, Neuchâtel, Switzerland, pp. 1–13 (2014)
33. Wolf, B., Loïc, M., Kuonen, P.: FriendComputing: organic application centric distributed computing. In: Proceedings of the Second International Workshop on Sustainable Ultrascale Computing Systems, NESUS 2015, Krakow, Poland, pp. 117–119 (2015)
34. Kuonen, P., Clement, V., Bapst, F.: Securing the grid using virtualization: the ViSaG model. In: Proceedings of the Fifth International Conference on Cloud Computing, GRIDs, and Virtualization, Cloud Computing 2014, Venice, Italy, pp. 49–54 (2014)
35. Enabling Mobile Augmented and Virtual Reality with 5G Networks. AT&T (2018). https://about.att.com/content/dam/snrdocs/Foundry%20ARVR%20Public%20Whitepaper.pdf. Accessed 29 Jan 2019
36. VR and AR pushing connectivity limits. Qualcomm Technologies, Inc. (2018). https://www.qualcomm.com/media/documents/files/vr-and-ar-pushing-connectivity-limits.pdf. Accessed 29 Jan 2019
37. Augmented and Virtual Reality: the First Wave of 5G Killer Apps. Qualcomm Technologies, Inc. (2017). https://www.qualcomm.com/media/documents/files/augmented-and-virtual-reality-the-first-wave-of-5g-killer-apps.pdf. Accessed 29 Jan 2019
38. Koivisto, M., et al.: Joint device positioning and clock synchronization in 5G ultra-dense networks. IEEE Trans. Wirel. Commun. **16**(5), 2866–2881 (2017)
39. Koivisto, M., Hakkarainen, A., Costa, M., Kela, P., Leppänen, K., Valkama, M.: High-efficiency device positioning and location-aware communications in dense 5G networks. IEEE Commun. Mag. **55**(8), 188–195 (2017)
40. Talvitie, J., Valkama, M., Destino, G., Wymeersch, H.: Novel algorithms for high-accuracy joint position and orientation estimation in 5G mmWave systems. In: Proceedings of IEEE Globecom Workshops, pp. 1–7 (2017)
41. Talvitie, J., Levanen, T., Koivisto, M., Pajukoski K., Renfors, M., Valkama, M.: Positioning of high-speed trains using 5G new radio synchronization signals. In: Proceedings of IEEE Wireless Communications and Networking Conference, WCNC 2018, pp. 1–6 (2018)
42. Rastorgueva-Foi, E., Costa, M., Koivisto, M., Talvitie, J., Leppänen, K., Valkama M.: Beam-based device positioning in mmWave 5G systems under orientation uncertainties. In: Proceedings of Asilomar Conference on Signals, Systems, and Computers (2018)
43. Witrisal, et al.: High-accuracy positioning for indoor applications: RFID, UWB, 5G, and beyond. In: Proceedings of 2016 IEEE International Conference on RFID, pp. 1–7 (2016)
44. Nam, S.Y., Joshi, G.P.: Unmanned aerial vehicle localization using distributed sensors. Int. J. Distrib. Sens. Netw. **13**(9), 1–8 (2017)
45. Zhang, P., Lu, J., Wang, Y., Wang, Q.: Cooperative localization in 5G networks: a survey. ICT Express **3**(1), 27–32 (2017)
46. Dammann, A., Raulefs, R., Zhang, S.: On prospects of positioning in 5G. In: Proceedings of 2015 IEEE International Conference on Communication Workshop (ICCW), pp. 1–7 (2015)
47. Tosic, M., Nikolic, Z., Nejkovic, Dimitrijevic, B., Milosevic, N.: Spectrum sensing coordination for FIRE LTE testbeds, Invited Paper. In: Proceedings of 2nd International Conference on Electrical, Electronic and Computing Engineering, IcETRAN 2015 (2015)

48. Boury-Brisset, A.C.: Managing semantic big data for intelligence. Proc. STIDS **2013**, 41–47 (2013)
49. Nikolic, Z., Tosic, M., Milosevic, N., Nejkovic, V., Jelenkovic, F.: Spectrum coordination for intelligent wireless Internet of Things networks. In: Proceedings of 2017 25th Telecommunication Forum (TELFOR), Belgrade, Serbia, pp. 1–7 (2017). https://doi.org/10.1109/telfor.2017.8249326
50. Nejkovic, V., Petrovic, N., Milosevic, N., Tosic, M.: The SCOR ontologies framework for robotics testbed. In: 2018 26th Telecommunication Forum (TELFOR), Belgrade, Serbia, pp. 1–4 (2018). https://doi.org/10.1109/telfor.2018.8611841
51. Petrovic, N., Nejkovic, V., Milosevic, N., Tosic, M.: A semantic framework for design-time RIoT device mission coordination. In: Proceedings of 2018 26th Telecommunication Forum (TELFOR), Belgrade, Serbia, pp. 1–4 (2018). https://doi.org/10.1109/telfor.2018.8611845
52. Jelenkovic, F., Tosic, M., Nejkovic, V.: Semantic driven code generation for networking testbed experimentation. Enterp. Inf. Syst. 1–17 (2018). https://doi.org/10.1080/17517575.2018.1509135
53. Redis. http://redis.io/topics/faq. Accessed 29 Jan 2019
54. Cois, C.A., Palko, T.: High-volume data collection and real time analytics using Redis. In: Proceedings of STRATA 2013. O'Reilly (2013)
55. Nakamura, S.: LINE Storage: Storing billions of rows in Sharded-Redis and HBase per Month. http://tech.naver.jp/blog/?p=1420
56. Apache Hadoop. http://hadoop.apache.org/. Accessed 29 Jan 2019
57. Apache HBase. http://hbase.apache.org/. Accessed 29 Jan 2019
58. Yao, Y., et al.: Online anomaly detection for sensor systems: a simple and efficient approach. Perform. Eval. **67**(11), 1059–1075 (2010)
59. Apache Jena. http://jena.apache.org/. Accessed 29 Jan 2019
60. OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/. Accessed 29 Jan 2019
61. Prud'Hommeaux, E., Seaborne, A.: SPARQL query language for RDF. W3C recommendation, 15 (2008)
62. Stocker, M., Shurpali, N., Taylor, K., Burba, G., Rönkkö, M., Kolehmainen M.: Emrooz: a scalable database for SSN observations. In: Proceedings of First Joint International Workshop on Semantic Sensor Networks and Terra Cognita, SSN-TC/OrdRing@ISWC, pp. 1–12 (2015)
63. Broekstra, J., Kampman, A., van Harmelen, F.: Sesame: a generic architecture for storing and querying RDF and RDF schema. In: Horrocks, I., Hendler, J. (eds.) ISWC 2002. LNCS, vol. 2342, pp. 54–68. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-48005-6_7
64. Todorovic, B., Rancic, S., Markovic, I., Mulalic, E., Ilic, V.: Named entity recognition and classification using context hidden Markov model. In: 2008, 9th Symposium on IEEE Proceedings of Neural Network Applications in Electrical Engineering (NEUREL), pp. 1–4 (2008)
65. Hermans, F., et al.: A lightweight approach to online detection and classification of interference in 802.15. 4-based sensor networks. ACM SIGBED Rev. **9**(3), 11–20 (2012)
66. Iyer, V., Hermans, F., Voigt, T.: Detecting and avoiding multiple sources of interference in the 2.4 GHz spectrum. In: Abdelzaher, T., Pereira, N., Tovar, E. (eds.) EWSN 2015. LNCS, vol. 8965, pp. 35–51. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15582-1_3
67. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
68. Ahmad, I., et al.: 5G security: analysis of threats and solutions. In: Proceedings of 2017 IEEE Conference 4 on Standards for Communications and Networking (CSCN), pp. 193–199 (2017)

69. Fang, D., Qian, Y., Hu, R.Q.: Security for 5G mobile wireless networks. IEEE Access **6**, 4850–4874 (2018)
70. Cho, S., et al.: System level simulation for 5G cellular communication systems. In: Proceedings of 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 296–299 (2017)
71. Mattisson, S.: Overview of 5G requirements and future wireless networks. In: Proceedings of ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference, pp. 1–6 (2017)
72. Nichita, M.V., et al.: 5G propagation: current solutions and future proposals. In: Proceedings of 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC), pp. 47–50 (2016)
73. Third Generation Partnership Project. http://www.3gpp.org/. Accessed 29 Jan 2019
74. Nagul, S.: A review on 5G modulation schemes and their comparisons for future wireless communications. In: Proceedings of 2018 Conference on Signal Processing and Communication Engineering Systems (SPACES), pp. 72–76 (2018)
75. Suthaharan, S.: Big data classification: problems and challenges in network intrusion prediction with machine learning. SIGMETRICS Perform. Eval. Rev. **41**, 70–73 (2014)
76. Casetti, C., et al.: Cognitive Network Management for 5G (white paper). In: 5GPPP Working Group on Network Management and QoS (2017). https://bscw.5g-ppp.eu/pub/bscw.cgi/d154625/NetworkManagement_WhitePaper_1.pdf. Accessed 29 Jan 2019
77. Al-Jarrah, O.Y., et al.: Efficient machine learning for big data: a review. Big Data Res. **2**, 87–93 (2015)
78. Ayoubi, S., et al.: Machine learning for cognitive network management. IEEE Commun. Mag. **56**(1), 158–165 (2018)
79. Imran, A., Zoha, A., Abu-Dayya, A.: Challenges in 5G: how to empower SON with big data for enabling 5G. IEEE Netw. **28**(6), 27–33 (2014)
80. Jakóbik, A.: Big data security. In: Pop, F., Kołodziej, J., Di Martino, B. (eds.) Resource Management for Big Data Platforms. CCN, pp. 241–261. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44881-7_12
81. Xie, J., et al.: A survey of machine learning techniques applied to software defined networking (SDN): research issues and challenges. IEEE Commun. Surv. Tutor. **1**, 1–39 (2018)
82. Abdallah, S., et al.: A network management framework for SDN. In: Proceedings of 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1–4 (2018)