Multicasting over Emerging 5G Networks: Challenges and Perspectives

Giuseppe Araniti, Massimo Condoluci, Pasquale Scopelliti, Antonella Molinaro, and Antonio Iera

Abstract

The number and variety of mobile multicast applications are growing at an unprecedented and unanticipated pace. Mobile network providers are in front of a dramatic increase in multicast traffic load, and this growth is forecasted to continue in fifth-generation (5G) networks. The major challenges come from the fact that multicast traffic not only targets groups of end-user devices; it also involves machine-type communications (MTC) for the Internet of Things (IoT). The increase in the MTC load, predicted for 5G, calls into question the effectiveness of the current multimedia broadcast multicast service (MBMS). The aim of this paper is to provide a survey of 5G challenges in the view of effective management of multicast applications, and to identify how to enhance the mobile network architecture to enable multicast applications in future 5G scenarios. By accounting for the presence of both human and machine-related traffic, strengths and weaknesses of the state-of–the-art achievements in multicasting are critically analyzed to provide guidelines for future research on 5G networks and more conscious design choices.

Introduction

The demand for multicast applications over cellular systems continues to grow rapidly [1]. As a consequence, multicasting will play a key role in emerging 5G networks, as outlined in white papers (e.g., from the NetWorld2020[1] technology platform), research projects, and standard documents from the 3rd Generation Partnership Project (3GPP) [2]. Indeed, multicasting represents a viable and effective solution to simultaneously convey data to a group of terminals through point-to-multipoint (PtM) communication, with positive consequences on the capacity and the spectrum efficiency of cellular systems. Both features are crucial for the deployment of 5G networks [3], as witnessed, for instance, in the recent METIS[2] and 5GNOW[3] European research projects.

Presently, video communication is considered as the "killer" *human-oriented multicast application*. Cisco has stated that the video traffic carried by mobile networks will reach 15 exabytes by 2019 (13 times larger than 2014). As a matter of fact, enhanced video services, e.g., ultra high definition (UHD), 4K and 3D videos, are becoming popular thanks to the quality of service (QoS) capabilities of Long Term Evolution (LTE) and beyond systems [1]. These services, together with the wide range of entertainment, interactive and real-time applications filling our daily lives, pave the way for future 5G human-oriented multicasting.

At the same time, network densification is triggered by the wide diffusion of a very large (and unpredictable) number of low-power devices supporting the deployment of the Internet of Things (IoT) [3] and exchanging small amounts of data using machine-type communications (MTC) [3].

The 5G multicast scenario becomes more complex when considering that small-cells, underlying the macro coverage, will be used to enhance the received signal strength levels associated with human services, and to increase the capacity in MTC scenarios, where devices are typically located in challenging positions (e.g., indoors, basements) [1, 3]. It becomes evident that the 3GPP Multimedia Broadcast Multicast Service (MBMS) [2] needs novel architectural and procedural definitions to meet the multifaceted constraints of the expected 5G multicast services.

Although multicasting is attracting an increasing amount of attention from a wide research community, many challenges still hinder its effective deployment in 5G networks. By surveying the relevant literature on this topic, it emerges that, until now, the primary research efforts have focused on approaches to boost the data rate performance (based on short-range links [4, 5], beamforming [12–14], network coding [10, 11], etc.) in the view of an improved quality of experience (QoE) of subscribers. Not enough attention has been given to the design of architectural and procedural solutions to meet the new challenges of multicasting in 5G networks, in which the same priority has to be given to both human-type and machine-type group services.

The objective of this paper is to identify and provide a critical evaluation of the enhancements required by MBMS to meet the constraints of 5G human-oriented and machine-oriented multicast applications. For this purpose, the paper is organized in such a way to first identify multicast application scenarios and related requirements and then point out enhancements necessary to extend MBMS in order to satisfy the identified requirements. Next, we scan the recent literature on 5G multicasting in order to highlight the key enabling technologies already in place to support 5G group-oriented services and the issues still open. Lessons learned and future research directions are finally identified and discussed.

Application Scenarios

The mobile market scenario for future 5G multicast applications is expected to be characterized by two types of services [1, 3]: the evolution of 4G applications tailored for human users; and the definition of novel machine-based services. It is thus of primary importance to analyze in

Giuseppe Araniti, Pasquale Scopelliti, Antonella Molinaro, and Antonio Iera are with the University Mediterranea of Reggio Calabria.

Massimo Condoluci is with King's College London.

[1] http://www.networld2020.eu

[2] METIS–Mobile and wireless communications Enablers for Twenty-twenty (2020) Information Society, https://www.metis2020.com

[3] 5GNOW–5th Generation Non-Orthogonal Waveforms for Asynchronous Signalling, http://www.5gnow.eu

Digital Object Identifier: 10.1109/MNET.2017.1600067NM

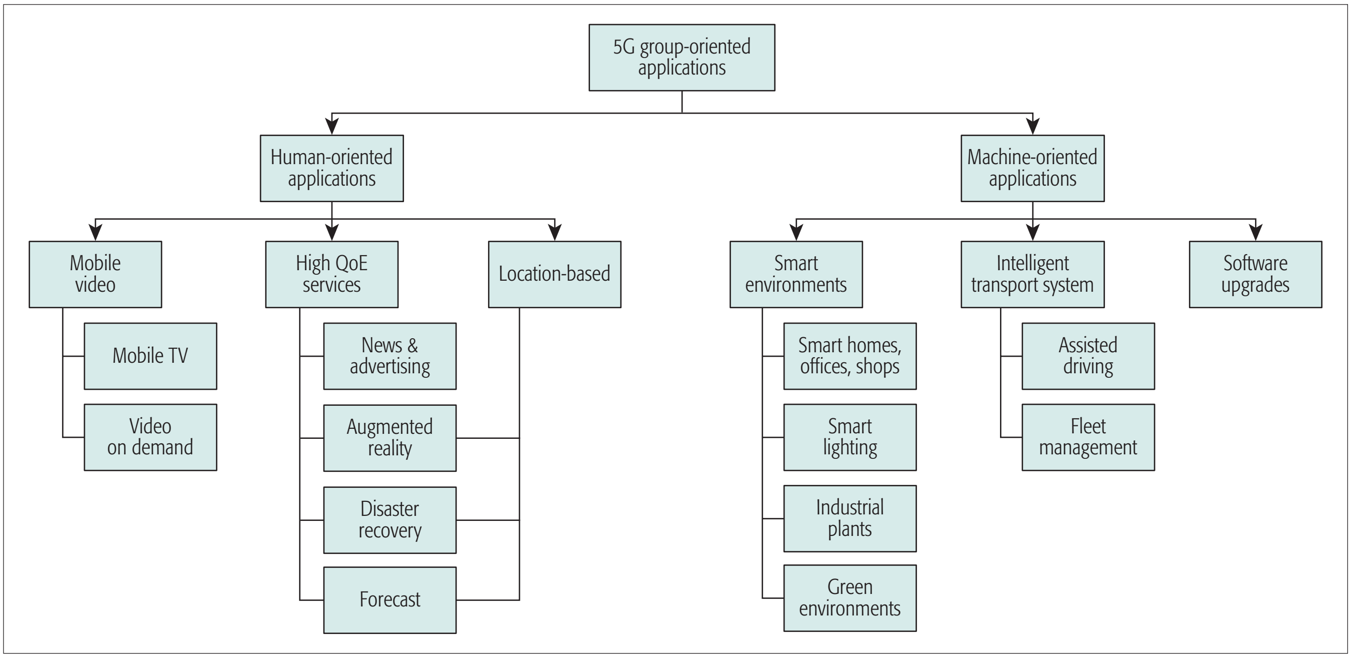

FIGURE 1. Application scenarios for multicasting over 5G systems.

depth these new services in order to highlight the requirements they dictate on the 5G network. A summary of these applications is given in Fig. 1 and will be discussed in the remainder of this section.

Human-Oriented Applications

Human-oriented applications represent the current trend of multicast services [1]. The future mobile market of human-oriented multicast applications can be envisioned as composed of the following service categories.

Mobile Video: Video downloading, video streaming, video conferences, sporting events, concerts and operas became very popular in recent years, and consequently, *group-oriented mobile TV* and *video on-demand* services [2] are expected to play a key role also in 5G systems. These classes of applications will be transmitted over 5G systems at ultra high definition (UHD) quality, and at the same time, enhanced 3D video capabilities.

The commercial success of such bandwidth-hungry applications is strongly tied to the development of effective solutions for resource allocation and management of co-existence with other (unicast and broadcast) services. This kind of application requires high data rates, low jitter, and connectivity everywhere, anytime, especially in mobility. Timing synchronization within the multicast group needs to be taken into account as well.

High-QoE Services: The improvement of users' QoE is a hot topic in the current literature [1], and this will be more important in several 5G multicast scenarios [3]. For instance, *news and advertising* applications can be enhanced by allowing a group of users to receive customized information according to their profiles (e.g., interests, hobbies, preferences). *Location-based* applications are a special class of high-QoE applications implying a service reception based on the user location. This implies the 5G concept of considering users being fully connected with the surrounding environment (as referred in a NetWorld2020's white paper[4] and in the METIS and 5GNOW EU projects). A first example of location-based applications are *augmented reality* multicast applications, especially conceived for commercial or tourist services. They allow users to receive additional information from the surrounding environment (e.g., visitors in a city/museum form a multicast group and receive interactive content related to the art work/rooms they are observing at that moment). Another target scenario for high-QoE location-based multicast applications is public safety for *disaster recovery*. In case of disasters (e.g., earthquake, fire, explosion), a group of users (both victims and rescuers) receive information of common utility to properly react to the emergency.

The classes of applications discussed above pose further challenges related to the management of users' position and profiles. This requires fine-grained user tracking mechanisms as well as effective procedures of group formation/joining and service announcement. In order to accommodate these new features, the current MBMS standard architecture needs to be upgraded. Furthermore, enhanced-QoE multicast applications ask for very low latency data transmission, high-reliability, and extended coverage to guarantee an adequate quality also to the terminals located in disadvantaged positions.

Machine-Oriented Applications

The benefits achieved by group-based, instead of unicast, communications toward machines are clear when considering the unprecedented huge number of connected sensors/machines expected for 5G systems [3]. A disruptive novelty introduced by group-oriented MTC is that the owner of the devices (i.e., a customer paying for cellular connectivity of all their equipment/appliances) can decide which device or group of devices to involve in the communication.

The capability of managing simultaneous multi-

[4] White paper on "5G: Challenges, Research Priorities, and Recommendations", available at http://www.networld2020.eu/sria-and-white-papers

Group-oriented applications		High data rates	Low-latency data delivery	Location-based group creation	Low-latency group creation	Customer-based group creation	Low-energy communication
Human-oriented services	Mobile video	✓	✓				
	High-QoE	✓		✓	✓		
Machine-oriented services	Smart environments		✓	✓		✓	✓
	Intelligent transport system		✓	✓	✓	✓	✓
	Software upgrades			✓		✓	✓

TABLE 1. Key requirements of 5G group-oriented applications.

ple MTC devices supports the delivery of different multicast machine-oriented applications such as those in the following subsections.

Smart Environments: The deployment of smart environments is a key goal of 5G/MTC platforms aiming at supporting stakeholders and users to make decisions based on real-time information in the view of significantly reducing costs, improving the quality of life, optimizing industrial processes, etc. [3]. In this context, multicast MTC applications can be beneficial for *smart homes/offices/shops*, when, for example, users out of their own homes send messages to a group of actuators to switch on/off electronic appliances (e.g., the heating/cooling system). Another case is the *smart lighting* application for homes, offices, or streets. For example, to save energy, lights on a mountain street can be turned on only when a car is getting close; thus, a group of lights can be switched on/off according to the movement of the cars. In a similar way, lights at home or in an office can be monitored and managed according to user needs. Furthermore, multicasting is mandatory for *smart industrial plants* to enable the efficient transmission of security control, warning, or management messages. For instance, in case of a problem in the industrial chain, the whole group of devices belonging to the assembly line could be stopped or re-organized to react to the critical event.

The above considered classes of applications can also be suitable to multicast *green environments*, where group-based management of sensors/actuators can reduce energy consumption (one of the most critical issues of 5G networks), thus increasing battery lifetime.

In conclusion, multicast applications related to smart environments require low-latency data transmission, low-energy communication, and location-based and customer-based group creation procedures. These requirements introduce tough challenges in the efficient and wise group formation (several classes of location-based applications) and in the reduction of the overhead for multicast transmission toward involved devices.

Intelligent Transport Systems: Vehicular applications developed in the last few years are claimed to be efficiently supported over 5G networks. Roads and vehicles will be equipped with sensors and tags to receive/transmit control/data messages [1, 3]. An example is *assisted driving*, where terminals involved in the same services (e.g., traffic management) or within the same area (e.g., cars close to the position of an accident) can be grouped to better disseminate data (e.g., traffic measurement, positions, speed) among interested vehicles. Similarly, *fleet management* applications can benefit from multicast transmissions.

Multicast transmissions for intelligent transport systems pose challenges in terms of the design of low-latency group formation/re-formation (also location-based) procedures, which are made more complex by the high speed of involved devices.

Software/Firmware Upgrade: Sensors, smartphones, and in general all smart devices, need software/firmware updates periodically or at the occurrence of specific events/dates. Smart devices could receive software/firmware upgrades either in case a new version is available or to fix bugs and add/change functionalities. In addition, software upgrades can also be location-based when, for instance, sensors installed in a given area receive upgrades to enhance/update their sensing capabilities (e.g., novel route directions on a street). In this case, the key challenge is related to group formation, which can be driven by the owner of sensors (who, for example, could be interested in sending data only to its own devices according to their location or functionalities/tasks). This means that not the network provider, but the sensors' owner should manage group formation. Therefore, the definition of effective customer-based group formation procedures becomes an issue.

The key requirements discussed in this section with reference to the different types of applications are summarized in Table 1. In the following, we will discuss how to meet such requirements by considering both architectural and data transmission points of view. To this aim, we first analyze the pros and cons of current group-oriented architectures and their possible evolution to meet 5G group-oriented requirements, and then we focus on the contributions in literature mainly covering data transmission aspects. We finally discuss the further enhancements needed to handle effective group-oriented applications in 5G systems.

Toward 5G Group-Oriented Network Architecture

The focus of this section is on the network architecture and procedures needed to handle multicast services. First, we present the current (4G) MBMS architecture, its design drivers and operation, and then we identify the major architectural and procedural changes needed to handle 5G group-oriented services in order to fit the requirements of Table 1. Emphasis will be given to machine-type services, which require more disruptive changes.

MBMS in a Nutshell

MBMS represents the reference standard architecture for multicast and broadcast service delivery in cellular systems [1, 2]. It specifies the network entities and the related interfaces as well as the procedures for supporting multicast services over 3GPP networks. The MBMS architecture includes:

- Broadcast multicast-service center (BM-SC), i.e., the source of multicast content that authorizes and initiates the MBMS bearer services and delivers MBMS data.
- MBMS-gateway (MBMS-GW), which accomplishes data content forwarding to the eNode-Bs (eNBs) involved in the MBMS session.
- Multicell/multicast coordination entity (MCE), which is in charge of session control signalling toward the involved eNBs.

More in detail, the BM-SC is in charge of providing *membership*, *session and transmission*, *service announcement*, and *security* functions that manage authorizations for MBMS subscribers. The BM-SC session and transmission function schedules MBMS session transmissions and retransmission and sends MBMS data. Through the service announcement function, the BM-SC is able to provide the user equipment (UE) with media descriptions, specifying the media to be delivered as part of an MBMS user service. The MBMS security function is used for distributing MBMS keys (key distribution function) to authorized UEs.

The MBMS-GW is located between the BMSC and eNBs, and its principal function is the sending/broadcasting of MBMS packets to each eNB transmitting the service. It allocates the IP multicast address to the eNBs involved in the delivery of MBMS traffic, and implements MBMS session control signaling (session start/update/stop) toward the E-UTRAN via the mobility management entity (MME).

The MCE is a logical entity with tasks of admission control and radio resource allocation to all eNBs, to decide suspension of MBMS session(s), to decide not to establish the radio bearer(s) of the new MBMS service(s) if the radio resources are not sufficient for the corresponding MBMS service(s). The MCE is involved in MBMS session control signaling. Moreover, an eNB is served by a single MCE

A summary of the functionalities of MBMS entities, its limitations and future enhancements can be found in Table 2.

Enhancing MBMS to 5G Multicast Services

The future scenario of 5G multicast services will be composed of heterogeneous environments, characterized by dissimilar communication ranges and capabilities, wherein an evolved MBMS manages different radio access and transmission technologies. In brief, wide area coverage is offered through geostationary satellites, managed by the satellite-eNodeB (S-eNB) located in the ground component, while macro eNBs provide group services in urban/suburban areas. Finally, small-cells (e.g., femto-cells, a.k.a. home-eNBs) and short-range (either 3GPP or non-3GPP) links enhance connectivity for indoor home/industrial services and extend the coverage of traditional macro-cells. The main architectural changes to the MBMS system are needed in order to manage multicast MTC.

	Node function and/or procedure	Limitations and/or future enhancement
BM-SC	• Service announcement function	• Extension for MTC missing
	• Membership function	• Extension for MTC missing
	• Security function	• Low processing capabilities of machines
MBMS-GW	• Data forwarding	• Heterogeneous eNB management
	• Session control signaling	• Signaling messages overloading • Signaling overhead
MCE	• Admission control and radio resource allocation	• Huge number of devices • MTC traffic with low energy requirements
	• Session control signaling	• Signaling messages overloading • MTC signaling

TABLE 2. Entities, related functionalities, and limitations of 4G MBMS.

Figure 2 shows an extended network architecture that leverages some enhancements to MBMS in order to support group-oriented MTC via small cells. The BM-SC must be enhanced to offer "customer-based" machine-oriented group services in addition to the traditional (human) services initiated by network providers. In the MTC case, in fact, only a set of MTC/IoT devices belonging to the same service and controlled through their own service capability server (SCS), could be interested in receiving data.

This implies that service announcement and membership functionalities should be tailored to deal with a pre-defined list of devices provided by the SCS. The reason is that machines cannot autonomously decide to join a multicast group; thus, membership has to be *customer-driven*, instead of provider-driven, as only the devices' owner is aware of which terminals belong to the MTC/MBMS service. The BM-SC should also be in charge of informing the devices involved in the group-oriented machine-type traffic, by triggering a *paging procedure* to enable MTC/IoT devices in *idle* mode to receive data. This aspect is highlighted in Fig. 3, where the differences between the legacy MBMS procedures and the enhanced MBMS procedures for MTC are illustrated.

The MBMS-GW will have to manage heterogeneous types of eNBs involved in the MBMS session, i.e., eNB(s)/HeNB(s)/S-eNBs. The issues relevant to this feature are heightened in 5G environments, where the number of involved base stations might be large, especially when considering small-cells. *Signaling* toward the mobile core network could be overloading, due to the huge number of involved eNBs.

At the same time, the MCE needs to implement more complex control signaling procedures to account for both human and machine-related group services, wherein the former are served with legacy MBMS control procedures while the latter ask for improved control procedures to support membership and paging functions. The current MBMS architecture does not provide any mechanism for *paging coordination* among different types of base stations, which is an issue raised by heterogeneous 5G networks.

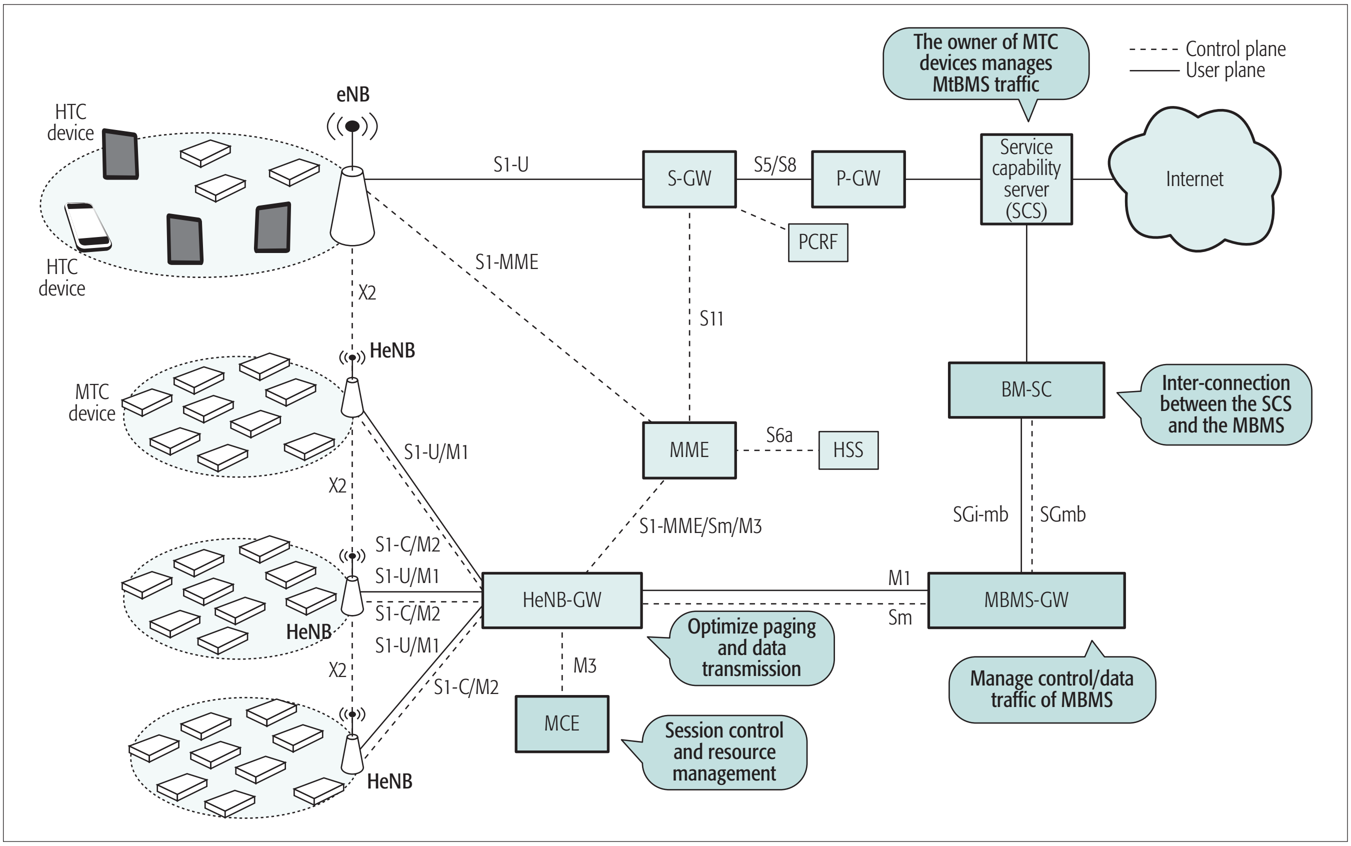

FIGURE 2. Enhanced MBMS architecture to support 5G group-oriented machine-type communications.

Toward 5G Group-Oriented Data Transmission: Preliminary Studies

The technology enhancements expected in 5G systems offer unprecedented potential for the delivery of multicast applications. In this section, recent achievements in data transmission for multicasting over cellular systems are surveyed.

Short-Range Enhanced Communications

Most of the literature on mobile multicasting has traditionally focused on boosting the data rates experienced by mobile users through the synergistic use of macro-cellular and short-range (both 3GPP and non-3GPP) links [5]. This approach can also assure a low data transmission delay. As for machines, the advantages of short-range communications also include prolonged terminal battery lifetime due to the reduced transmission power. In the remainder of this section, we will analyze the different improvements short-range communications bring to 5G multicasting.

The literature on this topic mainly focuses on the use of device-to-device (D2D) [6, 7] cellular links to *increase data rates* of multicast users. Researches in [4–9], for instance, proposed to serve only the portion of users with good channel conditions directly from the eNodeB, and use them as relay-entities to forward data toward the remaining receivers over short-range D2D links. The main issue in this scenario is cluster formation, i.e., the selection of the right amount of relaying devices to serve via the base station.

In this research area, the literature mainly addresses the impact of different cluster formation techniques on the multicast data rate. For instance, the main idea in [4] is to use intra-cluster D2D retransmissions to minimize the amount of needed resources, and this allows achieving about 40 percent gain in resource utilization when compared to legacy multicasting in cellular systems. This result shows the spectral efficiency improvement guaranteed by the joint exploitation of cellular and D2D links.

Energy-efficiency issues are considered in [8], where the authors focus on a two-stage cooperative multicast scheme to minimize the total transmission power without affecting the service coverage. Obtained results demonstrate that in a macro-cell about 80 percent of power consumption can be saved thanks to the wise use of D2D communications. Recently, the authors of [9] proposed a novel approach for D2D cluster formation, which encompasses the previous techniques. By assuming macrocell-driven synchronization of the transmitting relay nodes and the interest of all receiving devices in the same multicast content, they propose to allocate resources on the same frequency to all relay nodes for multicast data forwarding over D2D links. This reduces inter-cluster interference. In addition, they design a clustering scheme that minimizes the number of relays, thus reducing the *overall power consumption*.

A different family of works addresses the use of non-3GPP (e.g., Wi-Fi, Wi-Fi Direct) links to enhance the performance in macro-cells. The authors in [10] proposed a clustering algorithm that exploits macro-cellular and Wi-Fi Direct links to improve the session quality experienced

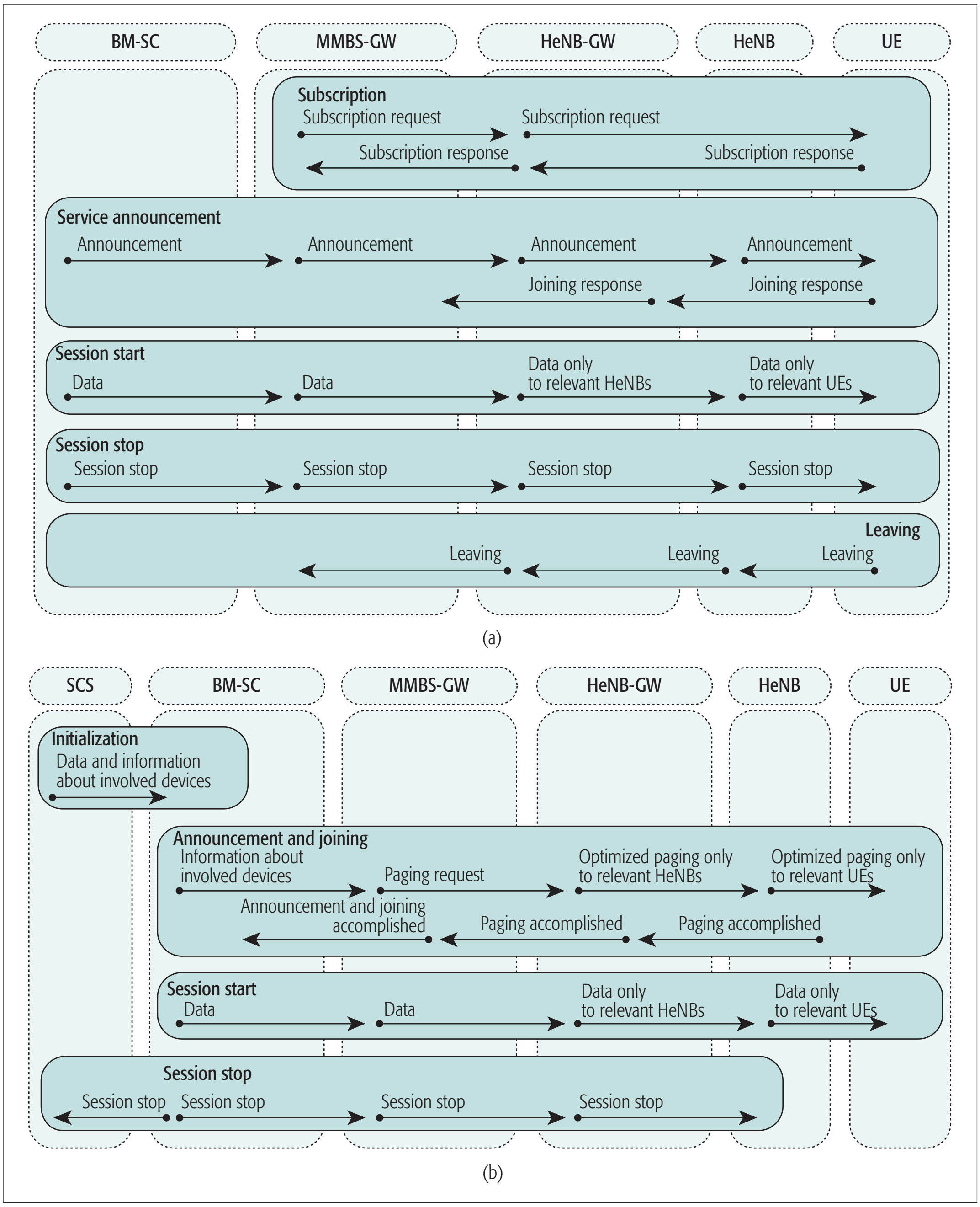

FIGURE 3. Enhanced MBMS vs. legacy MBMS procedures to support group-oriented services: a) legacy MBMS procedures; b) enhanced MBMS procedures for machine-type communications.

by multicast users by *reducing content delivery delay and energy consumption*. They also demonstrate that Wi-Fi links bring more beneficial effects than D2D links when the available spectrum is scarce.

A further concept of interest for multicasting enhancement over 5G networks is the social network group (SNG), where group members not only share their personal interests and keep in touch, but also share content. In this field, the authors in [11] introduced a cooperative multicast algorithm where a portion of SNG members download content of common interest through a cellular link and disseminate it among other group members through short-range ad-hoc wireless interfaces. The node(s) responsible for content downloading is dynamically chosen based on cellular link quality, residual battery level, and the amount of the content already downloaded by the members.

The MCE needs to implement more complex control signaling to account for both human and machine-related group services, wherein the former are served with legacy MBMS control procedures while the latter ask for improved control procedures to support membership and paging functions.

The literature review discussed above fits the requirements presented in the Table 1. For example, from the point of view of the mobile-video scenario, it is already demonstrated that short-range communication is suitable for providing high-data rate, low-latency, and low-energy communication.

Beamforming, i.e., the possibility of using directive antennas to achieve spatial diversity and throughput improvements, is also a technology that can boost the performance of multicasting over future 5G systems.

Macro/Small Cell Cooperation

Cooperation among macro and small cells is another key aspect that characterizes 5G systems, hence the interest in also evaluating cooperation in multicast scenarios. For instance, the authors in [12] introduce an enhanced architecture for multicast transmissions over heterogeneous cellular networks aimed at improving energy efficiency in multicast delivery. In this scenario, macro-cells provide wide-area coverage while small base stations (e.g., femto-cells) provide *high data rates* locally *in hotspot regions*. By analyzing user behavior, the idea is to multicast the data stream in the macro-cell region. Subsequent arriving users who (within a certain time window) request the same stream will immediately join the multicast group and the missing fraction of the stream will be served through small cells. It is worth noting that this solution is quite different w.r.t. the exploitation of D2D links, as no buffering is required on the user side. The main benefit of macro/small cell cooperation is that cooperative schemes outperform conventional unicast and multicast schemes in terms of *energy efficiency*. The cooperation between heterogeneous cells satisfies some of the requirements listed in Table 1, such as low-latency and low-energy communications. Although many works in the literature deal with this topic, 5G multicasting requirements of low-latency and low-energy need to be further investigated.

Network Coding

Network coding (where the packet decoded by a receiver is obtained as the combination of different information with different coding characteristics) has proven to be a further effective solution to enhance throughput and robustness of data transmission.

In [13], the authors proposed random linear network coding integrated with an unequal error protection (UEP) technique to improve the reliability of a layered multicast service. A layered service (e.g., H.264/SVC video) consists of a base layer, for basic service quality, and multiple enhancement layers, to improve the quality; with UEP, layers that are more important require a higher protection level. They also design a resource allocation framework that minimizes the number of coded packet transmissions needed to deliver service layers. Analyses on H.264/SVC video delivery over LTE-Advanced show that this approach *improves the service quality* (i.e., the number of consecutive layers recovered by a user) w.r.t. legacy PtM transmission by, at least, a factor of 1.35.

The use of network coding has also been considered in short-range D2D scenarios, as in [14]. Here, network coding is used to enable the transmission of different information coded into a single multicast packet destined to different receivers. A user-specific bit mapping algorithm applies different coding rates to different information before network coding is performed. Furthermore, a user-specific link adaptation scheme chooses an optimal modulation and coding scheme (MCS) for D2D multicast so that each user obtains their information from the same packet with different MCS according to the respective channel quality. This solution achieves a *throughput gain* ranging from 13 percent to 45 percent compared to conventional multicast transmissions.

Other preliminary works show that network-coded transmissions can improve spectrum re-use. In [15], for example, a new transmission system based on hierarchical spectrum re-use, named Cloud Transmission (CloudTxn), is proposed for terrestrial broadcasting (DVB-T) and can be properly adapted for PtM multimedia services. The idea is that two flows (A and B) share the same resources. Specifically, a (clock synchronized) data stream B is injected onto the same resources used by stream A, and transmitted with reduced power compared to stream A. This mechanism permits, for instance, the assignment of a more robust MCS to stream A (e.g., for mobile services in very noisy environments), and a more efficient MCS to stream B (e.g., to provide high quality HD and UHD services). It also *enhances spectrum efficiency and re-use*, which are key benefits expected in 5G systems.

Works presented above demonstrate that network coding techniques allow high throughput performance, hence these may be suitable for high data-rate human-oriented applications. Since network coding provides high data-rate performance, it is helpful to fit requirements of both mobile-video and High QoE applications.

Beamforming

Beamforming, i.e., the possibility of using directive antennas to achieve spatial diversity and throughput improvements, is also a technology that can boost the performance of multicasting over future 5G systems. Indeed, beamforming *improves the data rate* of users with the lowest channel gain without increasing the transmission power level, as for instance addressed in [15]. In detail, the authors in [16] present a low-complexity method that reduces the complexity of selecting the beamforming vector (i.e., the information needed by the antennas to make a directive transmission) and increases the throughput performance up to 52 percent w.r.t. conventional transmissions.

The focus of [17] is joint transmit beamforming and antenna selection in scenarios with multiple co-channel multicast groups. The algorithm presented in [17] finds, for each multicast group, the beamforming vector that minimizes the overall transmission power. A key goal achieved by this approach is the reduction in the number of antennas required to meet the quality constraints of multicast services.

Finally, adaptive beamforming for scalable video coding (SVC) scenarios is considered in [18], where the authors propose to schedule different SVC layers with different beams and MCSs while guaranteeing the respect of the QoE constraints of all multicast users. The increase in throughput performance, favored by beamforming, satisfies *high data rate* requirements of human-oriented applications, whereas directive antennas reach devices/machines with low channel quality, thus *extending the coverage*. Consequently, machine-oriented applications can also gain from the implementation of beamforming techniques. As discussed for network coding, beamforming is useful to fit high-data rate communications. Hence, mobile video and high-QoE applications could benefit from this technique.

Lessons Learned and Future Research Directions

Taking a cue from enabling features and emerging research trends outlined by the most recent literature, this section summarizes the lessons learned and discusses the open issues that still need to be thoroughly investigated in view of meeting the expected requirements of future 5G multicast systems and applications.

Lessons Learned

The technology enablers discussed earlier mainly focused on the radio-related aspects of multicast communications in order to improve the efficiency of data transmission. Indeed, the exploitation of short-range links, macro-cell and small-cell cooperation, network coding and beamforming brings key benefits to the data plane. In detail, by considering the requirements listed in Table 1, we can summarize the lessons learned from the literature as follows.

High data rates: Short-range communications support enhanced data rates thanks to the better channel conditions experienced by devices that are closer to each other w.r.t. the base station. In a similar way, small cells can be exploited to increase data rates in hotspot areas such as stadiums in case of events. In addition, data rates can be boosted through network coding techniques, which improve the robustness of data transmission. Finally, beamforming can boost data rates by exploiting directive antennas, thus improving the channel gains due to the increase in the received power by devices.

Low-latency data delivery: Short-range communications, macro/small cell cooperation, network coding and beamforming reduce latency in data delivery as they improve the data rates of data transmission from the base station to the devices. Nevertheless, the exploitation of short-range links introduces delays due to relaying operations, while network coding introduces delays to decode the received packets. These aspects did not receive enough attention in order to understand in which scenarios the final latency is effectively cut w.r.t. that experienced with legacy transmissions from the base station.

Location-based group creation: By exploiting the intrinsic feature of local communications, short-range links or small cells could be exploited to perform a location-based group creation. In this, only the devices in a limited area will receive the control information to join the multicast transmission, in contrast with the case when the joining procedure is performed through the macro-cell. Location-based services could be potentially offered by exploiting beamforming with a proper re-direction of beams. Nevertheless, the above-mentioned aspects have not yet been investigated. Finally, from a session management point of view, location-based group creation could be enabled by extending the functionalities of the BM-SC, with the joint exploitation of SCS in case of machine-type services.

Short-range communications as well as the exploitation of small cells could reduce the energy consumption of devices when receiving data thanks to the proximity nature of these communications.

Low-energy communications: Short-range communications as well as the exploitation of small cells could reduce the energy consumption of devices when receiving data thanks to the proximity nature of these communications.

From this brief synthesis, it is evident that a missing topic for the effective provisioning of 5G group-oriented applications is the control traffic needed to handle multicast services. Indeed, some of the requirements listed in Table 1 have not yet been addressed in the current literature, i.e., low-latency, location-based, and customer-based group formation. For this resaon, in the remainder of this Section we will focus our attention on future research directions dealing with control traffic management.

Future Research Directions

The support of machine-based services especially raises several challenges to be solved.

A first issue is related to low-latency group creation. MTC devices need to be switched on (from *idle* to *connected mode*) in order to receive multicast data. Consequently, an issue for the customer is to identify the machines in the multicast group and to page them before data transmission begins. Group paging procedures in 5G systems are expected to be designed with the aim of simultaneously switching a large number of devices into connected mode, with further positive effects on the *control overhead* and the *latency* of group-based machine applications. Indeed, MTC applications may require transmissions of a few bytes each. Thus, minimizing the amount of control information exchanged between the network and the devices becomes an issue of utmost importance, not only from a latency point of view. Besides, due to the machines' low processing capabilities, MTC/IoT communications suffer from *security issues*, since it is not easy to implement security functions, e.g. key exchange procedures. These aspects need to be considered as they affect the group formation delay.

As emphasized in [4–10], a large number of related works in the literature exploiting 3GPP (i.e., D2D) and non-3GPP (e.g., Wi-Fi) short-range links assume that the unit responsible for resource allocation (generally, the base station) is "somehow" aware of the channel conditions of short-range links among group members. In such scenarios particular interest in future research should be given to *control messaging* (for the purpose of nearby device discovery, cluster formation/updating, etc.). In practical systems, the way the base station becomes aware must be specified. It is straigthforward to think that each terminal measures (or estimates) the link quality with nearby devices and transmits this information to the resource management unit. Nonetheless, effective solutions in the case of large multicast groups still need to be properly investigated and designed to avoid overloading or bottleneck

Enabling features	Expected benefits	Open issues
Short-range direct links	• High data rates • High spectrum efficiency (spectrum re-use in case of 3GPP D2D links) • Offloading of cellular spectrum (in case of non-3GPP links) • Coverage extension • Reduced energy consumption	• D2D cluster formation (device discovery, control messaging among group members) • Non-3GPP cluster formation (device discovery, control messaging among group members) • New criteria for relay nodes selection (i.e., buffer status, battery level) • Interference induced by D2D links on non-multicast cellular users • Cluster re-formation procedures (e.g., triggered by device mobility, or running out-of-energy)
Network coding	• High data rates • High robustness • High spectrum efficiency	• Complexity at the transmitter and receiver sides • Latency • Robustness to packet losses • Control overhead
Beamforming	• High data rates • High spectrum efficiency (spatial diversity)	• Complexity • Control overhead • Fast re-computation in case of mobile users • Jointly use directive antennas and MIMO
Machine-type communication	• MTC group-services • New business opportunities for Telco • Simplification of group management procedures (e.g., announcement, join)	• Extension of legacy MBMS architecture • Definition of machine-oriented MBMS procedures • Definition of multicast paging procedures • Control overhead • Improvements of resource efficiency and system capacity

TABLE 3. Main research directions and challenges for 5G multicasting.

The higher the number of involved devices, the higher the complexity of the algorithms, and this affects latency in terms of both group creation and data delivery.

effects at the base station in 5G scenarios. In addition, control messaging among interested devices needs to be properly defined, and their performance assessed by considering their impact on MTC traffic and devices in terms of latency and energy-consumption. The above mentioned issues are exacerbated in the case of non-3GPP short-range links, which require the effective solutions to manage control and billing messaging between, e.g., LTE and trusted/non-trusted Wi-Fi interfaces. The highlighted problems do not yet find answers in the literature. Thus, there is an urgency to focus greater attention on them to enable effective short-range solutions for enhanced multicasting.

A further research need in terms of low energy communications deals with the procedure of *relay selection*. Indeed, while current works mainly focus on channel conditions, additional information about buffer state, residual battery levels or position of nodes could be relevant to enhance 5G relay selection/updating procedures. In particular, relay selection based on device position could potentially enable location-based multicast services. Hence, a constantly increasing number of devices with constrained power and memory capabilities challenges multicast IoT services.

The higher the number of involved devices, the higher the *complexity* of the algorithms, and this affects latency in terms of both group creation and data delivrey. This is still linked to the above discussed limited computational power of machine-devices. More powerful devices, instead, face the problem of increasing energy consumption. In general, the complexity represents the most challenging issue to be solved in 5G systems. Although the use of network coding has emerged as a promising approach to improve the reliability of multicast transmissions [12] without additional resource requirements, the complexity problem is exacerbated by the expected large amount of MTC devices with reduced computational capabilities. In addition, the computational cost for beamform calculation at the base station increases data delivery latency. This calls for low-complexity near-optimal solutions valid in a wide range of environments and users' locations and propagation conditions.

From a more general point of view, an effective low-latency (as well as a customer-based) group creation can be achieved through a re-design of joining procedures achieved by extending the functionalities of BM-SC and MBMS-GW entities (Table 3).

Conclusions

In this paper, the main challenges and future directions of multicasting over future 5G systems are outlined. We critically analyzed the current state of the art on multicasting, when it is enhanced by key 5G technologies, such as device-to-device communications, network coding and beamforming. The surveyed literature provides interesting results, though these are mostly aimed at improving the data rates experienced by multicast users and spectrum utilization. There is a lack of contributions in the area of group-based machine-type communications that require important enhancements to the current MBMS architecture. To fill this gap, we present research needs and future trends by underlining the open research challenges and the issues that still need to be addressed in the view of effective human-based and machine-based multicast implementations over 5G systems.

References

[1] D. Lecompte and F. Gabin, "Evolved Multimedia Broadcast/Multicast Service (eMBMS) in LTE-Advanced: Overview and Rel-11 Enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, Nov. 2012, pp. 68–74.

[2] 3GPP, "TS 23.246, Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description, Rel. 13," Mar. 2015.

[3] F. Boccardi *et al.*, "Five Disruptive Technology Directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 74, 80.

[4] B. Zhou *et al.*, "Intracluster Device-to-Device Relay Algorithm with Optimal Resource Utilization," *IEEE Trans. Vehic. Tech.*, vol. 62, no. 5, June 2013, pp. 2315–26.

[5] L.Song *et al.*, *Wireless Device-to-Device Comunications and Networks*, Cambridge University Press, 2015.

[6] J. Liu *et al.*, "Device-to-Device Communication in LTE-Advanced Networks: A Survey," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 4, 4th Quarter 2015, pp. 1923–40.

[7] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-Smartphone: Realizing Multihop Device-to-Device Communications," *IEEE Commun. Mag.*, vol. 52, no. 4, Apr. 2014, pp. 56–65.

[8] Y. Zhou *et al.*, "Two-Stage Cooperative Multicast Transmission with Optimized Power Consumption and Guaranteed Coverage," *IEEE JSAC*, vol. 32, no. 2, Feb. 2014, pp. 274–84.

[9] L. Militano *et al.*, "Single Frequency-Based Device-to-Device-Enhanced Video Delivery for Evolved Multimedia Broadcast and Multicast Services," *IEEE Trans. Broadcasting,* vol. 61, no. 2, June 2015, pp. 263–78.

[10] L. Militano *et al.*, "When D2D Communication Improves Group Oriented Services in Beyond 4G Networks," *Wireless Networks*, Nov. 2014, pp. 1–15.

[11] J.-B. Seo, T. Kwon, and V. Leung, "Social Groupcasting Algorithm for Wireless Cellular Multicast Services," *IEEE Commun. Lett.*, vol. 17, no. 1, Jan. 2013, pp. 47–50.

[12] X. Zhang *et al.*, "Energy-Efficient Multimedia Transmissions through Base Station Cooperation over Heterogeneous Cellular Networks Exploiting User Behavior," *IEEE Wireless Commun.*, vol. 21, no. 4, Aug. 2014, pp. 54–61.

[13] A. Tassi, I. Chatzigeorgiou, and D. Vukobratovic, "Resource-Allocation Frameworks for Network-Coded Layered Multimedia Multicast Services," *IEEE JSAC*, vol. 33, no. 2, Feb. 2015, pp. 141–55.

[14] T. Zhou *et al.*, "User-Specific Link Adaptation Scheme for Device-to-Device Network Coding Multicast," *IET Commun.*, vol. 9, no. 3, Feb. 2015, pp. 367–74.

[15] Y. Wu *et al.*, "Cloud Transmission: A New Spectrum-Reuse Friendly Digital Terrestrial Broadcasting Transmission System," *IEEE Trans. Broadcasting*, vol. 58, no. 3, Sept. 2012, pp. 329–37.

[16] W. Lee *et al.*, "A New Beamforming Design for Multicast Systems," *IEEE Trans. Vehic. Tech.*, vol. 62, no. 8, Oct. 2013, pp. 4093–97.

[17] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint Multicast Beamforming and Antenna Selection," *IEEE Trans. Signal Proc.*, vol. 61, no. 10, May 2013, pp. 2660–74.

[18] H. Park *et al.*, "Adaptive Beam Scheduling for Scalable Video Multicast in Wireless Networks," *Electronics Lett.*, vol. 50, n. 16, July 2014, pp. 1167–69.

There is a lack of contributions in the area of group-based machine-type communications that require important enhancements to the current MBMS architecture. To fill this gap, we present research needs and future trends by underlining open research challenges and issues that still need to be addressed.

Biographies

Giuseppe Araniti (araniti@unirc.it) is an assistant professor of telecommunications at the University Mediterranea of Reggio Calabria, Italy. From the same university he received the Laurea (2000) and the Ph.D. degree (2004) in electronic engineering. His major area of research includes personal communications systems, enhanced wireless and satellite systems, traffic and radio resource management, multicast and broadcast services, device-to-device (D2D) and machine-type communications (M2M/MTC) over 5G cellular networks. He is a senior member of IEEE.

Massimo Condoluci (massimo.condoluci@kcl.ac.uk) received the B.Sc. and M.Sc. degrees in telecommunications engineering in 2008 and 2011, respectively, from the University Mediterranea of Reggio Calabria, Italy. From the same university he received the Ph.D. degree in information technology in 2016. He is currently a research associate at the Centre for Telecommunications Research (CTR), King's College London, UK. His current research interests include network and function design, virtualization, softwarization, machine-type communications and multicasting over 5G wireless networks.

Pasquale Scopelliti (pasquale.scopelliti@unirc.it) received the M.Sc. in telecommunications engineering in 2013 from the University Mediterranea of Reggio Calabria. He is currently a Ph.D. Student at the same university. His current research is focused on cellular systems, radio resource management, multicast and broadcast services over 5G cellular networks, and heterogeneous Networks.

Antonella Molinaro (antonella.molinaro@unirc.it) is an associate professor of telecommunications at the University Mediterranea of Reggio Calabria, Italy. She was an assistant professor with the University of Messina, Italy, from 1998 to 2001; with the University of Calabria, Cosenza, Italy, from 2001 to 2004; and with the Polytechnic of Milan, Italy, from 1997 to 1998, with a research contract. She was with Telesoft, Rome, from 1992 to 1993, and at Siemens, Munich (DE), from 1994 to 1995, with a European fellowship contract. Her current research interests include vehicular networking and future Internet architectures.

Antonio Iera (antonio.iera@unirc.it) graduated in computer engineering from the University of Calabria, Italy, in 1991, and received a master diploma in information technology from CEFRIEL/Politecnico di Milano, Italy, in 1992 and a Ph.D. degree from the University of Calabria in 1996. Since 1997 he has been with the University of Reggio Calabria and currently holds the position of full professor of telecommunications and director of the Laboratory for Advanced Research into Telecommunication Systems (www.arts.unirc.it). He has been an IEEE senior member since 2007. His research interests include next generation mobile and wireless systems, RFID systems, and Internet of Things.