

Group Communications in Narrowband-IoT: Architecture, Procedures, and Evaluation

Galini Tsoukaneri, Massimo Condoluci, *Member, IEEE*, Toktam Mahmoodi, *Senior Member, IEEE*, Mischa Dohler, *Fellow, IEEE*, Mahesh K. Marina, *Senior Member, IEEE*

Abstract—Narrowband-IoT (NB-IoT) has been released by 3GPP to provide extended coverage and low energy consumption for low-cost machine-type devices. Requiring only a reasonably low-cost hardware update to the already deployed Long Term Evolution (LTE) base stations and being compatible with current core network and enhanced core solutions that aim to reduce the battery consumption and minimize the signalling, NB-IoT deployments are quickly increasing, making NB-IoT a dominating technology for low-power wide area (LPWA) networks. To this aim, in this paper we focus on group communications (i.e., multicast) in NB-IoT to efficiently support the transmission of firmware, software, task updates or commands towards a large set of devices. We discuss the architectural and procedural enhancements needed to support the unique features of group communications in machine-type environments, such as customer-driven group formation. We also extend the NB-IoT frame to include a channel for multicast transmissions. Finally, we propose two transmission strategies for multicast content delivery and evaluate their performance considering the impact on the downlink background traffic and the channel occupancy.

Index Terms—NB-IoT, MTC, MBMS, Group communications, Multicast.

I. INTRODUCTION

Narrowband-Internet of Things (NB-IoT) is a new cellular network technology that was designed for improved power consumption, spectrum efficiency and deep coverage. It is primarily targeting low cost IoT and Machine Type Communication (MTC) devices with infrequent connections, that are expected to operate for 10 years or more [1], [2] (e.g. sensors, meters). Based on the expectation for low traffic by these devices NB-IoT operates on just 180kHz of system bandwidth within the LTE spectrum, and thus provides very limited communication resources. Therefore, it is even more imperative than in LTE, that these resources are used efficiently. However, the majority of works so far (e.g. [3]–[6]) have focused only on the uplink direction, as this constitutes the majority of the expected traffic.

The long lifetime necessitates that devices can receive software updates to remain secure and up-to-date [2]. Furthermore, they are likely to be receiving commands to execute tasks.

G. Tsoukaneri and M. K. Marina are with the School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK. e-mail: G.Tsoukaneri@sms.ed.ac.uk, mahesh@ed.ac.uk.

M. Condoluci, T. Mahmoodi, and M. Dohler are with the Department of Informatics, King's College London, London WC2B 4BG, UK. e-mail: {massimo.condoluci, toktam.mahmoodi, mischa.dohler}@kcl.ac.uk.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

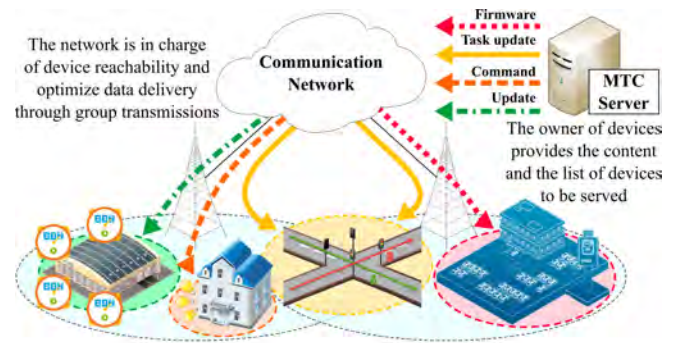


Fig. 1. MTC devices gathered in different groups, each one receiving a specific content.

As such, efficient use of the downlink direction is equally important. The large amount of involved devices running the same software and performing the same tasks, leads us towards *group communications*, i.e. multiple devices receiving the common content simultaneously to save network resources, instead of being served one-by-one (unicast).

In current LTE networks, group services are provided with the Multimedia Broadcast Multicast Service (MBMS) [7]. MBMS follows a subscription-based approach, where the mobile operator periodically announces all available services (e.g. firmware update) in the Multicast Control Channel (MCCH) that devices can subscribe to. Each device has to monitor the periodic service announcements for relevant services, and when these occur, it subscribes to the evolved NodeB (eNB). Then, after setting up the appropriate communication bearers, it has to keep monitoring the MCCH for the exact time that the service will be broadcasted. The Single Cell - Point to Multipoint (SC-PTM) framework is a similar approach that extends MBMS to provide group communications within a single cell. More recently, it was also standardized in NB-IoT [8].

While this subscription-based model was successful in LTE for applications such as video-on-demand, it is inefficient in terms of resource utilization and energy consumption, and it is poorly suited for NB-IoT. The major drawback of the current MBMS (and by extension SC-PTM) is that it requires resources from the data transmission channel for the periodic service announcement, even when there are no currently available services to announce. This resource wastefulness is even worse for NB-IoT, considering that it already operates on very limited resources. Additionally, IoT devices would need to continuously monitor the service announcements, even

though services such as firmware updates are rare and far in-between. This wastes energy, as the device needs to wake up outside of its Discontinuous Reception cycle (DRX), and has a great impact on the battery life of the device.

Our main contribution is a novel set of enhancements to MBMS to address both resource utilization and energy consumption. To do so, we do not follow the subscription-based model as in SC-PTM. Instead, we page devices **on-demand** whenever there is a relevant service, foregoing the need for periodic service announcements and monitoring. Our approach is more flexible, allowing for dynamic creation of device groups (based on device make, performed tasks etc.) that will receive each service, instead of statically selecting at manufacture time which services the device will need. The basis of our approach is that the service provider (e.g. manufacturer, owner) presents the network operator with a list of devices that should receive the service, in addition to the service content itself. The network operator can then page the listed devices directly, so that they can receive the multicast transmission (Fig. 1).

To demonstrate the efficiency of our approach over SC-PTM, we conducted a thorough experimental evaluation using a custom simulator that implements the downlink specifications of NB-IoT. In particular, we measured the delay that each approach incurs to the existing unicast traffic, the delivery time of a firmware update as well as the total time devices spent receiving control information (uptime), and show that our approach outperforms SC-PTM in all cases. Finally, we investigated two different multicast transmission scheduling strategies for group communications.

The rest of the paper is organized as follows. Section II is an overview of NB-IoT, MBMS and SC-PTM. Section III presents our approach from a procedural and architectural point of view. Section IV details our experimental setup and presents our results. Section V discusses future work, and section VI provides our conclusive remarks.

II. BACKGROUND & RELATED WORK

A. Narrowband-Internet of Things (NB-IoT)

NB-IoT is designed to co-exist with LTE [9] and can be deployed in three operation modes depending on the LTE band used: (i) guard-band, (ii) in-band and (iii) stand-alone. Regardless of the operation mode, NB-IoT requires a minimum bandwidth of 180 kHz for both the downlink and uplink, i.e. one Physical Resource Block (PRB) split into 12 subcarriers of 15 kHz each. To facilitate interoperability with LTE deployments, NB-IoT uses the same frame structure and numerology with LTE [10]. Frames are 10ms long and are composed of 10 subframes of 1ms each. Subframes are split into two slots of 0.5ms, each composed of 7 Orthogonal Frequency Division Multiplexed (OFDM) symbols and a normal cyclic prefix (CP) is used in all subframes.

The smallest transmission unit is the Resource Element (RE) and is composed of one OFDM symbol and one subcarrier. To preserve the interoperability with LTE in the in-band mode, a control region of two or three first OFDM symbols in each subframe is allocated to the Physical Downlink Control

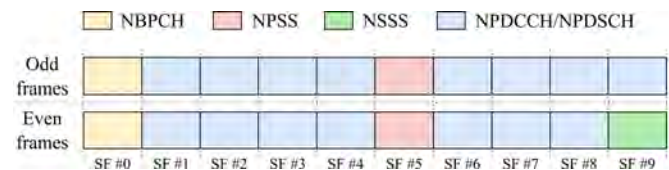


Fig. 2. Structure of NB-IoT frames for odd and even frames.

Channel (PDCCH) of LTE. The control region size is signalled in the Narrowband System Information Block 1 (SIB1-NB) and for the guard-band and stand-alone modes is equal to 0. In the uplink Quadrature Phase Shift Keying (QPSK) or Binary Phase Shift Keying (BPSK) modulation can be used, while in the downlink only QPSK is allowed.

NB-IoT provides a coverage enhancement of 20dBm compared to legacy LTE with the aim to reach devices in signal challenged locations. Towards this end, transmission repetitions are exploited with a maximum value of 128 and 2048 for uplink and downlink respectively. Furthermore, 3 coverage enhancement (CE) levels are defined to cope with various radio conditions, each with different numbers of repetitions.

To provide low-complexity services, NB-IoT introduces a limited set of signals and channels compared to LTE [11]. As the focus of the paper is on the downlink, we now summarize the three new signals and three new channels of the downlink in NB-IoT.

Signals: The *Narrowband Reference Signal (NRS)* is similar to the LTE's Cell Reference Signal (CRS) and provides phase reference for channel demodulation. It is transmitted in all subframes regardless of whether actual data is also being transmitted. The *Narrowband Primary Synchronisation Signal (NPSS)* is used for synchronization, and is transmitted in subframe #5 of every frame (10ms). Finally, the *Narrowband Secondary Synchronisation Signal (NSSS)* is also used for downlink synchronization along with the NPSS, but it is transmitted in subframe #9 of every two frames (20ms).

Channels: The *Narrowband Physical Broadcast Channel (NPBCH)* is transmitted in subframe #0 and carries the Narrowband Master Information Block (MIB-NB). The *Narrowband Physical Downlink Control Channel (NPDCCH)* carries control and scheduling information both for the downlink and the uplink. It also carries Hybrid Automatic Repeat Request (HARQ) and random access response (RAR) messages as well as paging information for the devices. Lastly, the *Narrowband Physical Downlink Shared Channel (NPDSCH)* carries the SIBs-NB that carry important information regarding the configuration of the cell, as well as data towards the devices. NPDSCH also has the same structure as the NPDCCH in terms of NRS, CRS and control region [12]. The subframes eligible to carry NPDCCH or NPDSCH are #1-#4, #6-#8 in all frames and #9 in odd frames. Each of these subframes can carry one of the two channels and the exact allocation is done by means of scheduling. The NB-IoT downlink frame is depicted in Fig. 2. More details on NB-IoT can be found in [1], [9], [10], [13].

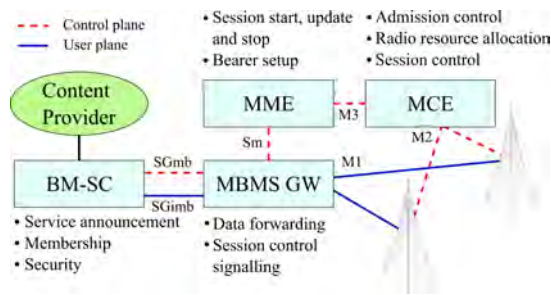


Fig. 3. MBMS architecture in the EPC with main functionalities for each entity.

B. Multimedia Broadcast Multicast Service (MBMS)

MBMS is a 3GPP standard to support point-to-multipoint services [14] by means of either *single-* or *multi-cell* (a.k.a. single-frequency) transmissions. It was first introduced in Release 6 but has significantly changed throughout the following releases to provide efficient transmissions of multicast content. MBMS over LTE/LTE-A is usually referred to as evolved MBMS (eMBMS), which also supports single-frequency transmissions. In order to handle the devices' subscriptions, the co-ordination of the eNBs that participate in the transmission and the delivery of the multicast data MBMS introduces new physical entities in the core network (Fig. 3). In the remainder of this Section the main features of MBMS are described. More details can be found in [7], [15], [16].

Channels: MBMS introduces two new logical channels, the *Multicast Transport Channel (MTCH)* and *Multicast Control Channel (MCCH)* for data and control respectively. To transmit multicast content, one or more PDSCH subframes are allocated to MBMS. Both channels can be mapped in the same MBMS subframe which may contain one MCCH and one or more MTCHs. At the MAC layer, MTCH(s) and MCCH are multiplexed in the *Multicast Channel (MCH)*, which is carried by the *Physical Multicast Channel (PMCH)* in the physical layer.

Procedures: The MBMS framework consists of eight different steps in order to deliver multicast content.

- *Subscription:* a user establishes an agreement with the content provider to receive MBMS services.
- *Service Announcement:* used by the network to inform the devices about the offered services.
- *Joining:* allows a device to join a multicast group before or during the MBMS service delivery.
- *MBMS Notification:* the network informs the devices about an imminent/ongoing service.
- *Session Start:* the network reserves the required resources (i.e., Evolved Packet Core (EPC) and radio bearers) for the service.
- *Data Transfer:* the period of time during which MBMS data is being transmitted.
- *Session stop:* the network releases the allocated resources.
- *Leaving:* allows user to leave a multicast group.

Operation: MBMS is subscription-based standard where the network operator periodically announces the available services and their status in the form of control information in the MCCH. Devices indicate which services they are interested in and receive a group id that uniquely identify the service. They

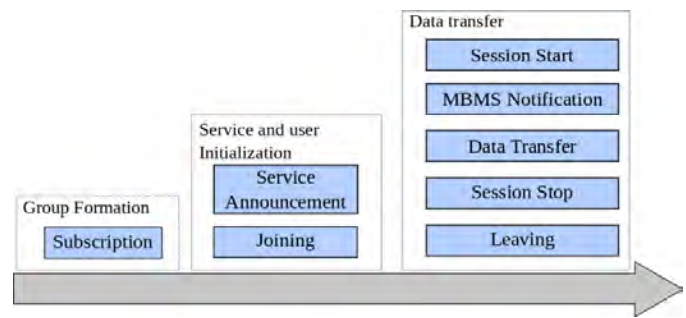


Fig. 4. Procedures for the reception of a MBMS service.

then need to monitor the MCCH for notifications regarding the start of their service and the setup of the communication bearers. Devices can join a service before it starts or while it is ongoing. Upon termination of the transmission all bearers and resources are released and the sessions is terminated. The session stop does not indicate the termination of the service. Subscribed devices still need to periodically monitor the MCCH for information regarding future sessions of their service. In contrast, when a device chooses to leave a service it discards the group ID and can then stop monitoring the MCCH for control information. In MBMS, the subscription, joining and leaving are performed on a per-user basis while the rest of the procedures are performed on a per-MBMS service. The MBMS procedures are depicted in Fig. 4

Transmission modes: MBMS supports two transmission modes depending on how many eNBs are involved in the transmission. In the *single-cell* mode each eNB transmits MBMS data to only the devices in its cell, thus transmission parameters are adapted independently from each eNB. In the *multi-cell* mode a set of eNBs simultaneously transmit the same MBMS content. The involved eNBs compose a single frequency network (SFN) and they are tightly synchronized in time and frequency to transmit the same data in the same frames/subframes and frequency resources. In the multi-cell approach only the extended cyclic prefix may be used. By extending the cyclic prefix of OFDM symbols to avoid inter-symbol interference, devices perceive the signals from different eNBs as a single transmission thus increasing the signal-to-interference ratio. However, the use of the extended cyclic prefix limits the number of resources for data transmission and thus reduces the spectral efficiency.

C. Single Cell - Point to Multipoint (SC-PTM)

Several works have highlighted the need for multicast transmission in 5G and NB-IoT networks (e.g. [10], [13], [15], [17]). To that end, 3GPP introduced the *Single Cell - Point To Multipoint (SC-PTM)* framework to provide multicast transmissions which was recently standardized also for NB-IoT [8].

SC-PTM is a complimentary bearer service to the existing MBMS. It uses two logical channels, the *Single Cell - Multicast Control Channel (SC-MCCH)* and the *Single Cell - Multicast Traffic Channel (SC-MTCH)* for control and data respectively. However, in contrast to the previous implemen-

tation of MBMS where the MCCH and MTCH occupied a whole subframe, the new channels are scheduled in NPDCCH and transmitted in the NPDSCH multiplexed with the unicast traffic. The control information is transmitted periodically with a scheduling period specified by the network operator [8]. The multicast data is scheduled using a group Radio Network Temporary Identifier (G-RNTI) that uniquely identifies the service. To deliver the data *Single Cell - Multimedia Radio Bearer (SC-MRB)* is used, which is set up before the session start and accessed by all devices in the group.

As SC-PTM inherits the subscription-based architecture of MBMS, it also requires periodic announcements and monitoring of SC-MCCH, so it is poorly fitted for the resource limited NB-IoT. For the broadcasting of control information a number of NPDSCH resources needs to be allocated to the SC-MCCH on regular intervals regardless of whether ongoing transmissions exist. Especially in the cases of firmware updates or task commands, the services need to be available all the time as new devices need to be able to subscribe at any given time and there is no indication as to when the next session will start. The constant availability requires constant use of NPDSCH resources for the transmission of the SC-MCCH, but in NB-IoT where the available resources are limited, this may result in severe degradation of the system's performance as precious resources are wasted for control information of pending multicast services. Furthermore, devices that have subscribed to SC-PTM services need to periodically monitor the SC-MCCH for information regarding their services. The subscription and monitoring is done on a per-service basis and on top of their own DRX thus increasing the energy consumption of the devices.

For NB-IoT multicast services we can assume that devices subscribe to at least one service after their first power on (e.g. to receive updates from the vendor). Hence, the number of concurrently available services should be at least equal to the number of types/makes/models of the devices present in the cell as even different models of the same manufacturer may require different updates. In SC-PTM the maximum number of concurrent services is 64 [8]. However, due to the variability of the devices, the applications they run and the fact that the services need to be available all the time, the number of different services will be much larger. Furthermore, from the devices' perspective, deciding which services a device should subscribe to can be very challenging as it is the device manufacturer, the application provider or the device owner that decides which updates or commands should be delivered to which devices and the different multicast groups may be created on the fly. For all the aforementioned reasons SC-PTM is not well suited for the types of devices and applications targeted by NB-IoT.

III. PROPOSED FRAMEWORK FOR NB-IOT

As seen in the previous Section, MBMS is based on a complex procedure that introduces high latency and energy consumption for devices, as well as a direct interaction with the devices. To align with NB-IoT's goals of low complexity and long battery life, several enhancements need to be introduced to MBMS. In this section we discuss how group

communications can be supported in NB-IoT. To this end, we first present a set of modified procedures of MBMS able to cope with machine-type use cases, with particular focus on multicast scenarios. We then present new mechanisms for grouping, paging and delivering content to devices involved in a group communication. Finally, we discuss how to efficiently support multicast transmissions within a NB-IoT frame and present two different transmission methods that try to reduce the impact of multicast traffic on unicast traffic.

A. Group Communications for NB-IoT

Our proposed enhancements to the MBMS standard (Fig. 5) focus on the multicast transmission mode as it allows a co-ordination entity (e.g., owner of MTC devices, device manufacturer) to decide which devices must receive a specific content.

The *subscription* assumes that a co-ordination entity decides which devices belong to a certain multicast group. It is thus independently performed (e.g. by the device owner) and the network is provided with a list of devices for each group through the MTC server.

The service announcement has been removed, as the concept of service availability is not applicable to MTC traffic in NB-IoT anymore as devices will be receiving content decided and provided by the co-ordination entity.

The *joining* procedure is different from the legacy MBMS. Instead of waiting for a device to join a group, in our proposed joining procedure the network informs the devices about the multicast group(s) they belong to.

The MBMS notification procedure has also been removed to reduce the signalling towards the devices, and the session start procedure is used to inform the devices about an imminent service. The *session start* procedure is triggered when the MTC server provides the multicast content for a given group. Before the session start the network establishes the EPC bearers and activates existing multicast bearer in the Radio Access Network (RAN) (Sec. III-B). Once all bearers are setup and active the data transmission can start. After the content has been transmitted, the network triggers the *session stop* to release the bearers previously created in the EPC and deactivates the bearers in the RAN.

From an architecture point of view, the core network needs to be able to receive the list of devices from the device manufacturer or owner for a certain group store the information related to the membership of devices to a given group. Generally, the core network would implement enhanced policies for managing a group session for both the control and data traffic, while considering the energy consumption of the devices and the control traffic overhead.

B. Bearer Setup & Paging

In our proposal devices do not monitor service announcements but are instead paged to receive multicast data. Furthermore, communication bearers are required for the delivery of the data. However, bearer setting/activation for group communications needs to be managed in an efficient way to reduce the energy consumption of the devices as well as the

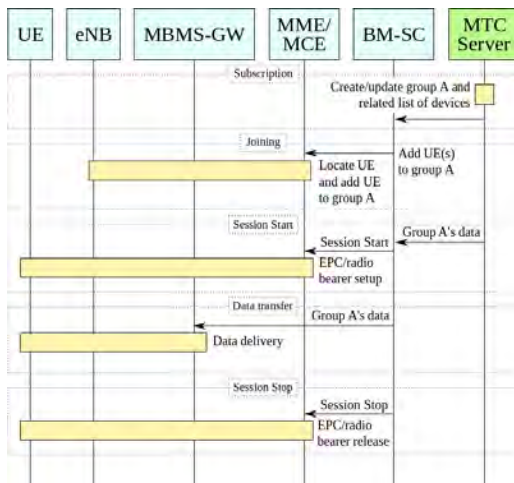


Fig. 5. Proposed enhancements to the MBMS standard.

control overhead. Therefore, we propose a combined *service activation & joining* procedure where NB-IoT devices set up a generic MBMS bearer (or join an existing one if already available) according to their preferred Quality of Service (QoS). Devices perform this step after their first session start process, where the eNB can adapt the transmission to the QoS supported by the devices and their bearers [18]. At the session stop process, the eNB releases the multicast bearers in the EPC but keeps the RAN multicast bearers in a idle status in order to be re-used in the future. NB-IoT devices perform the multicast bearer setup only once and skip the random access process in each subsequent paging, thus decreasing the latency and the energy consumption when receiving future multicast content.

With the multicast bearers setup and waiting, the devices need only to be paged when there is multicast data. However, MTC and in particular NB-IoT devices that are supposed to receive the same multicast data may have different DRX cycles, thus different paging occasions. It is inefficient to assign an additional paging occasion for the group(s) as the devices will waste energy in monitoring more than one paging occasion. This is particularly inefficient considering that many group communications (e.g., firmware updates) have intervals ranging from weeks to months.

Driven by [17], we propose a *grouping on-the-fly* scheme, where the device manufacturer or owner provides the network with the list of devices uniquely identified by their International Mobile Subscriber Identities (IMSI) to receive multicast content. For a certain group but the network manages the grouping of devices in a efficient way. While the owner of devices creates the groups without considering the DRX cycles, the network is aware of these values. As a consequence, the network can logically split the devices into subgroups based on their DRX cycles and perform paging accordingly. Their periodicities can also be taken into account for data delivery, i.e., syncing data transmission with the transmission of the uplink reports (in this case, the device wakes up only once to transmit uplink data and then it does not need to be waken up again for multicast delivery).

Paged devices skip the random access as the bearers already exist. In essence, they are instructed to activate their MBMS

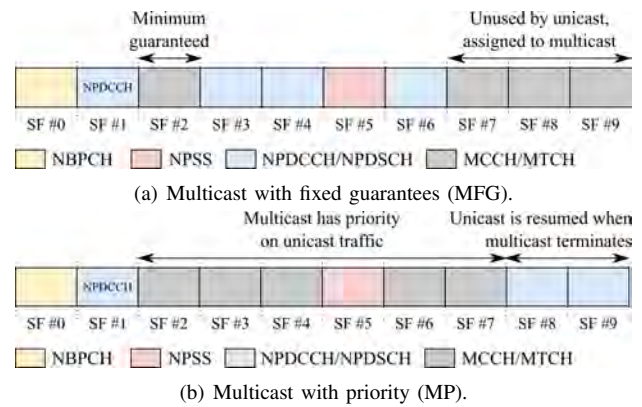


Fig. 6. Proposed transmission schemes applied to odd frames of NB-IoT.

bearer, after which they can immediately start receiving the multicast data. As there is only one bearer per group, the number of multicast bearers is not expected to be a significant burden to the eNB. This approach can be further optimized by allowing one bearer per device owner which will be used for groups/contents.

C. Proposed Transmission Strategies

In legacy LTE networks, multicast transmission is accomplished by allocating all PRBs of selected subframes to MBMS. However, NB-IoT is deployed on a single PRB thus significantly limiting the number of available resources. To efficiently support group communications in NB-IoT it is important to take into account the scarce resource availability. The limitation of resources becomes even more challenging if we consider the presence of unicast downlink traffic in the background, and transmission schemes for multicast traffic in NB-IoT must not significantly affect the existing background traffic. To this end, we propose two methods for multicast data transmission in NB-IoT, *Multicast with fixed guarantee (MFG)* and *Multicast with priority (MP)* (Fig. 6). A performance evaluation of the two strategies is provided in Section IV.

Multicast with fixed guarantee (MFG): The aim of this strategy is to minimize the impact of multicast traffic on unicast. This is achieved by giving priority to unicast traffic and transmitting multicast content only when there are available resources. However, as the unicast traffic is generally uniformly distributed in time, always giving it priority might result in having very limited resources for multicast. This will lead to increased delays for the multicast traffic and increased energy consumption for devices, since they will require longer time to receive data and might be paged sooner than their periodicities to receive outstanding multicast packet, whenever there is resource availability.

To overcome this problem, we propose *multicast with fixed guarantee (MFG)*, which is based on the idea of providing a minimum number of guaranteed resources per frame that can be used for multicast traffic, leaving the remaining resources for background unicast traffic. The minimum number of guaranteed resources for the multicast traffic applies to all frames until the end of the multicast transmission (including repetitions). Control and data information is carried using

logical control and traffic channels respectively and are delivered using the allocated resources. The amount of reserved resources is network operator specific and it can be decided based on the observed average traffic or past information on the same periods of time¹. The resources allocated to multicast traffic as well as the exact scheduling of the control and data information are signalled in the SIB20-NB, similarly to SC-PTM, and are guaranteed as long as there exist multicast data for transmission. However, in contrast to the SC-PTM where the SIB20-NB carries information regarding the SC-MCCH scheduling, modification periods etc., in our approach the SIB20-NB only informs the devices about where the control information and multicast data is. Furthermore, the control information only indicates how the multicast data should be decoded. In other times, all resources can be used for background traffic. An example of resource allocation is depicted in Fig. 6(a), where the minimum guarantee is equal to 1 subframe. It is worth noticing that resources which are not used for unicast transmissions might be scheduled for multicast.

Multicast with priority (MP): While MFG limits the impact of multicast over unicast traffic, it also limits the efficiency of multicast delivery as multicast traffic will only receive a limited amount of resources. This can increase the reception latency and by extension the energy consumption of the devices involved in the group communication.

To solve this issue, we further propose the *multicast with priority (MP)* strategy that aims to deliver the multicast content as quickly as possible, thus lowering the latency and the energy consumption. In MP, the network gives priority to multicast data by allocating all the NPDSCH subframes to the multicast transmission. The only exception is the transmission of the SIBs-NB. When the NB-SIBs need to be transmitted the required resources are not allocated to multicast. Unicast traffic is thus paused and then resumed after the end of multicast session. Similarly to the MFG approach, the same control and data channels are used and the resource allocation and scheduling of them are signalled in the SIB20-NB. An example of a frame allocation is depicted in Fig. 6(b).

The MP approach is suitable for infrequent multicast data, i.e., when the amount of multicast sessions is limited or when multicast sessions are sparse. Although it impacts the unicast traffic more compared to MFG, it only delays data reception for devices expecting to receive unicast data during the multicast. However, it is worth highlighting that the amount of suspended unicast transmissions due to the presence of multicast data is limited compared to the number of devices receiving multicast data, i.e. MP introduces delays for a smaller set of devices compared to the MFG approach. Moreover, delaying a unicast transmission might not involve higher energy consumption for the device, as the network can buffer the downlink packets and page the device to resume its reception after the multicast transmission is finished.

¹By exploiting the periodic nature of NB-IoT devices it is possible to predict when resources will be needed for unicast transmission and therefore choose to transmit multicast data in times that the expected unicast traffic is low.

IV. EXPERIMENTAL EVALUATION & RESULTS

A. Experimental Setup

To assess the impact of multicast traffic in the current NB-IoT technology and the performance of the proposed schemes, we conducted a thorough experimental evaluation using a custom simulator in Matlab, that implements the downlink specifications of NB-IoT. In our experiments we examine the behaviour of our approach (using the MFG and MP scheduling strategies), and compare it against SC-PTM.

For comparison, we measure the delivery time, the NPDSCH occupancy and the device uptime for monitoring control information. The *delivery time* is measured differently for the background and multicast traffic. For the background traffic, it is the time elapsed from the moment the eNB receives an ACK from the MTC server (hence, different for each device) to the moment the ACK is received by the device (including all repetitions). For the multicast traffic, it is the time elapsed from the moment the eNB receives the content (hence, the same instant for all devices involved) to the moment all devices receive the content (including all repetitions). For the *NPDSCH occupancy*, we measure the percentage of resources the NPDSCH used, over the overall availability in a frame. Finally, *uptime* is the average time that a device is awake. Since the uptime caused for data transmission is unavoidable and exactly the same over all approaches, we only measure the uptime due to monitoring control information and paging.

Our experiments were performed under the presence of typical NB-IoT background traffic (e.g. NB-SIBs, downlink Random Access messages, application ACKs), as found in [19]. Specifically, we simulated a 5 MHz LTE cell with an in-band NB-IoT deployment and 55000 NB-IoT devices. For each device we generated a valid DRX cycle, as outlined in [20]. For multicast content, we transmitted 1MB of data, which we believe to be representative of a typical firmware update for NB-IoT devices. Similar to the approach taken currently by 3GPP [21], we assumed the use of an Application Layer Forward Error Correction (AL-FEC) scheme based on fountain raptor codes, which increases the effective size of the multicast content to 1.2MB, due to the fountain raptor codes overhead.

In terms of network settings, we assume one downlink acknowledgement (ACK) of 142 bytes (45B packet size, 5B Internet Protocol (IP) header, 8B Packet Data Convergence Protocol (PDCP) header, 8B Radio Link Control (RLC) header, 16B Medium Access Control (MAC) header) per uplink transmission. Following the results of [19], we considered that ACKs follow the same distribution as uplink transmissions. Subframes #1 and #6 were allocated to NPDCCH, while the remaining subframes were allocated to NPDSCH. We assess all scenarios using both normal and extended CP for single- and multi-cell deployments respectively. Extended CP needs to be used when multiple eNBs form a SFN to decrease the signal interference. However, using extended CP in the NB-IoT subframes requires that the same subframes of the LTE deployment also have extended CP, thus reducing the overall system performance. Degrading the system performance is

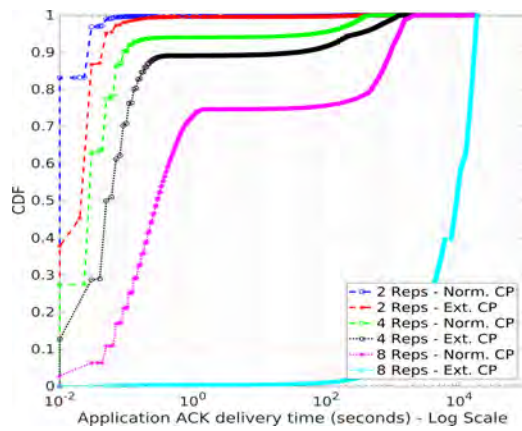


Fig. 7. **Unicast baseline ACK delivery time:** In normal operation the delivery time from a few hundreds of ms for 2 repetitions and normal CP to a few hundreds of seconds for 8 repetitions and extended CP.

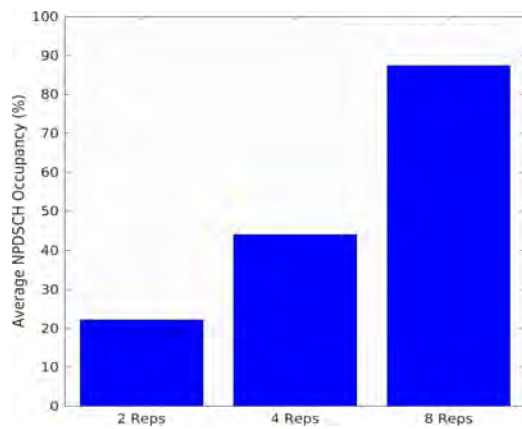


Fig. 8. **Unicast baseline NPDSCH occupancy:** the occupancy is about 20% with 2 repetitions and increases up to 90% with 8 repetitions.

of course undesirable, but it is worth exploring the system behavior in such cases, since SFN for MBMS is currently in use. Finally, we assume that devices remain awake for 1ms every time they monitor either the SC-MCCH or the paging channel.

B. Results

Unicast baseline: Initially, we wanted to assess the generic performance of the system when only background traffic exists, as it will allow us to understand the impact of multicast traffic on unicast in the other scenarios. Fig. 7 depicts the application ACK delivery time. On average, the delivery time varies from a few hundreds of ms up to a few seconds. The maximum delivery time is in the order of hundreds of seconds and it is obtained when 8 repetitions are used. The average NPDSCH occupancy (Fig. 8) is drastically affected by the presence of repetitions. The occupancy is of about 20% when 2 repetitions are used, and it increases up to 45% with 4 repetitions. When using 8 repetitions, the occupancy reaches a value of about 90%. Experiments with higher numbers of repetitions resulted in around 100% NPDSCH occupancy. Hence, we decided to use 2, 4 and 8 repetitions for the rest of the evaluation to better understand (i) the impact of multicast

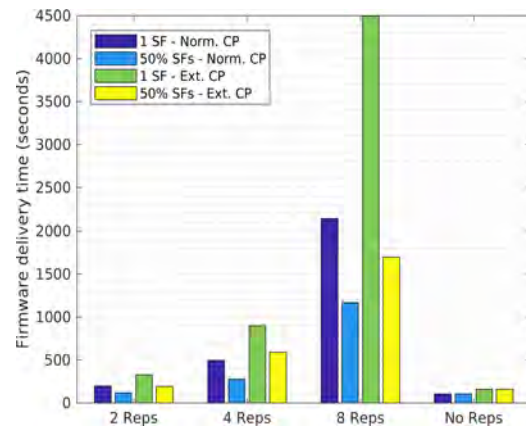


Fig. 9. **MFG firmware delivery time:** The maximum delivery time is around 4500s with 8 repetitions with 1 SF guaranteed and extended CP, while less than 500 seconds are required if 2 repetitions are used, regardless of the CP used.

transmissions on the NPDSCH occupancy and the (ii) benefits of our proposed strategies in reducing that occupancy.

MFG: In this scenario we considered one multicast transmission with 1, 2, 4 and 8 repetitions, in order to assess their impact on the firmware delivery time. We can see that the maximum delivery time is around 4500s with 8 repetitions for the case 1 SF guaranteed with extended CP (Fig. 9). We can also observe that the performance of firmware delivery deteriorates in a multi-cell scenario when the extended CP is used. In terms of impact on the unicast traffic, results are not depicted since giving priority to the unicast traffic does not drastically affect the performance in the case of single-cell scenario. The highest impact applies to the multi-cell scenario with 8 repetitions, since the use of extended CP reduces the amount of available resources and increases the application ACK delivery time.

We can also notice that the NPDSCH occupancy is not drastically affected compared to the unicast baseline when 1 SF is guaranteed for multicast (Fig. 10). The occupancy for the case of 50% of resources guaranteed shows an increase of ~8%, ~18%, and ~10% with 2, 4, and 8 repetitions respectively.

MP: In this experiment we evaluate the performance of the MP scheduling and its impact on the background traffic. Our results show that the firmware delivery time decreases for the case with 8 repetitions down to ~540s with normal CP and ~780s for the extended CP (Fig. 11). This testifies the benefits that the MP approach introduces for multicast traffic. At the same time, the ACK delivery time is almost similar to that obtained in the unicast baseline (Fig. 12). This denotes that the MP introduces delays only to a very limited number of unicast devices, i.e., those needing to receive data in the time interval of multicast traffic delivery, but it does not drastically affect the overall performance of unicast traffic. An interesting result is the NPDSCH occupancy (Fig. 13), as there are small differences compared to Fig. 8. Taking the latency into account these results show that the overall impact of MP approach is limited.

SC-PTM: In terms of firmware delivery time, we see that

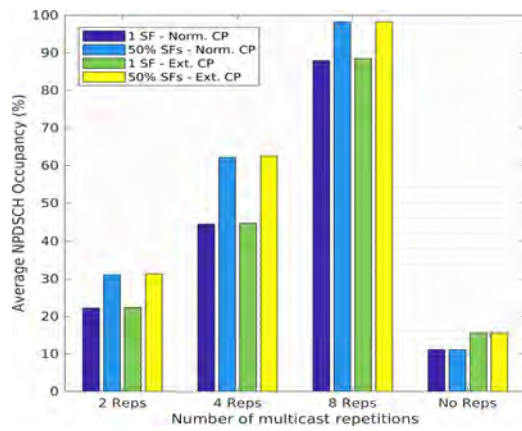


Fig. 10. **MFG NPDSCH occupancy:** The NPDSCH occupancy is not drastically affected when 1 SF is guaranteed, but increases when 50% of the resources are given to multicast.

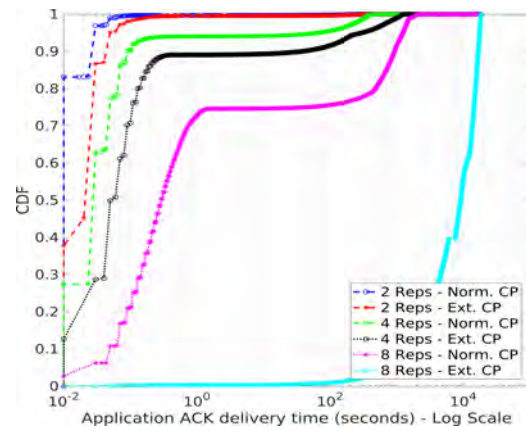


Fig. 12. **MP ACK delivery time:** The ACK delivery time is similar to the baseline indicating that the impact on the background traffic is minimal.

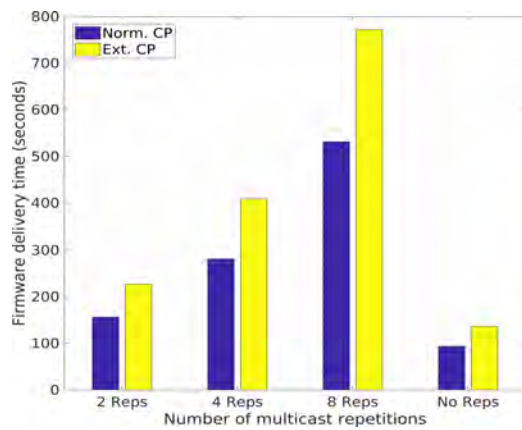


Fig. 11. **MP firmware delivery time:** As expected, the firmware delivery time is lower compared to MFG even when 8 repetitions are used.

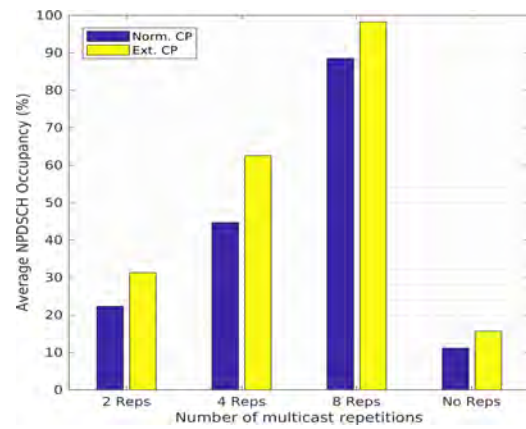


Fig. 13. **MP NPDSCH occupancy:** The NPDSCH occupancy remains in the same levels as in the unicast baseline.

our approach with the MFG scheduling strategy outperforms SC-PTM, even though it prioritizes the background traffic over the multicast transmissions (Fig. 14). This is expected as both approaches utilise all available resources not used by background traffic, but we also give a minimum guarantee of resources per frame for multicast data. When we use the MP strategy, which prioritizes multicast traffic, the margin over SC-PTM only grows. Furthermore, we do so while having a minimal impact on the background traffic (Fig. 12).

The device uptime in SC-PTM is heavily dependent on the scheduling period values of SC-MCCH. In our experiments, we considered three different scheduling periods that cover the range specified by 3GPP [8]: 2, 50 and 200 radio frames (RFs). For simplicity we assumed that devices were subscribed to only one service. Note that subscribing to more than one services would increase the uptime of the device even further, so this is the best case scenario for SC-PTM. We used a value of 1ms for the monitoring instance of both the SC-MCCH and the paging channel (as in our method).

We can see that the uptime over a period of a day is significantly larger for SC-PTM (Fig. 15). This is expected as devices under SC-PTM need to monitor both the paging channel and SC-MCCH. The uptime, of course, is directly

linked to the energy consumption of the device, and this can add up significantly over a span of 10 years.

We also compared the NPDSCH occupancy dedicated to multicast data and control information. For fairness we used the same amount of resources between SC-PTM and our proposal and the same number of repetitions. Note however, that our method guarantees a minimum number of resources while in SC-PTM the multicast content transmission is depended on the resource availability. Our approach essentially only uses the resources needed for data transmission and decoding information. Comparatively, SC-PTM uses significantly more resources, even when considering the extreme case of a 200 RFs scheduling period (Fig. 16). Of course such an extreme case comes at the expense of significantly increased delivery time (Fig. 14). When striking a balance between occupancy and delivery time (50 RFs), SC-PTM uses 2 times more resources than our method when 8 repetitions are used, or 4 times more resources with 2 repetitions. It is apparent that the subscription-based scheme of SC-PTM, wastes a significant amount of resources just to transmit control information, when these resources could efficiently be used for actual data transmission (multicast or unicast).

Finally, we demonstrate the behavior of each approach as the background traffic increases. For simplicity, we only

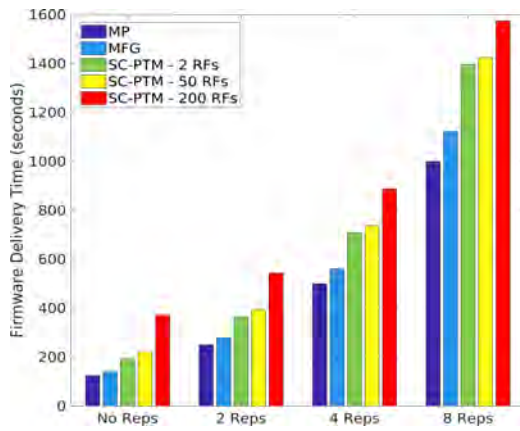


Fig. 14. **SC-PTM firmware delivery time:** We can see that both our proposed scheduling strategies outperform SC-PTM.

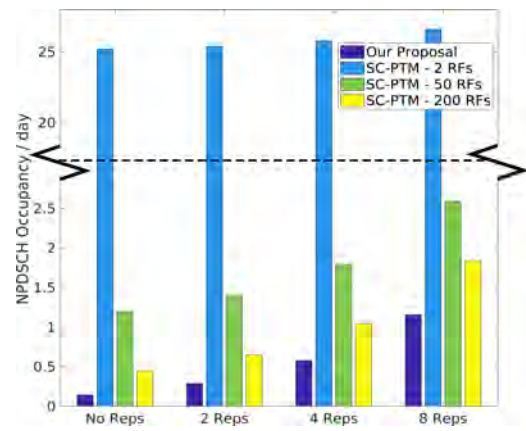


Fig. 16. **SC-PTM NPDSCH occupancy for control information:** Our proposal consistently uses less resources while SC-PTM uses significantly more resources, even when considering the extreme case of a 200 RFs scheduling period. Please note that the figure is split in two parts for better clarity.

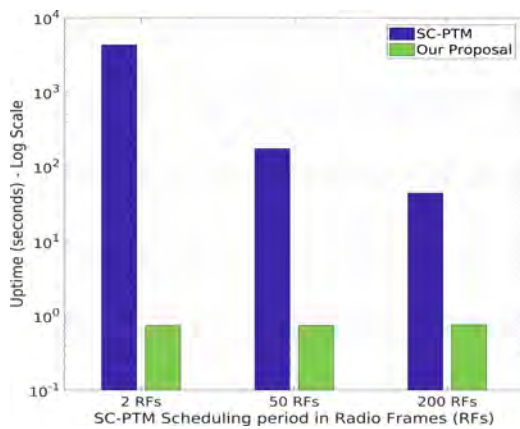


Fig. 15. **SC-PTM Uptime:** The uptime over a period of a day is significantly larger for SC-PTM compared to our proposal, as devices under SC-PTM need to monitor both the paging channel and SC-MCCH.

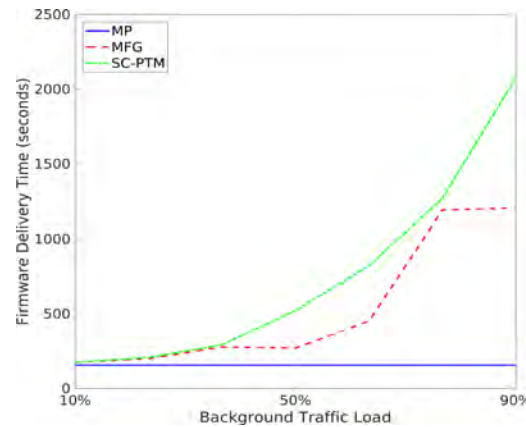


Fig. 17. **Firmware Delivery Time with different background traffic loads:** The benefits of the minimum guarantee of MFG are apparent under high background load conditions (90%), where it is able to deliver the multicast content faster.

present one set of parameters (normal CP with 2 repetitions, and 1 SF minimum guarantee for MFG), but the same trend can be observed in all cases. With respect to the firmware delivery time, we can see that our method (for both MP and MFG scheduling strategies) scales better than SC-PTM as the load increases (Fig. 17). This can be explained by the fact that SC-PTM needs to use part of the available resources for multicast control information, which results in fewer available resources for the multicast transmission itself.

What is important to note, however, is that the reduced firmware delivery time does not come at the expense of the background traffic. In fact, our method produces equal or better ACK average delivery times in all cases (Fig. 18). This is true even for the MP scheduling strategy, that prioritizes the multicast transmissions, as we are able to fully utilize all available resources for data transmission (either multicast or unicast).

V. FUTURE WORK

Efficient mechanisms to manage multicast bearers in the RAN and efficient sub-grouping of devices (for either paging or data delivery purposes) are two major points that are out of

the scope of this paper, but we strongly believe will improve the overall performance of group communications in NB-IoT.

Efficient bearer grouping at the time of the random access and attach process can have significant benefits on the usage of RAN resources and reduce the impact of multicast traffic on background traffic in the downlink (i.e., fewer bearers for multicast traffic thus more resources for background traffic). However, grouping dissimilar devices together in terms of QoS and transmission parameters may result in efficient resource utilization (i.e., inefficient transmission parameters even if only device does not have a good channel) and increased energy consumption.

Efficient sub-grouping of devices needs to be properly investigated. A large number of sub-groups reduces the efficiency of both paging and data delivery that is contradictory to the whole idea of group communications. It also increases the impact on the background traffic, especially in the MP case. At the same time, having one group for all devices means that each device will have to wait until every other device has also been paged before receiving any data, thus increasing the

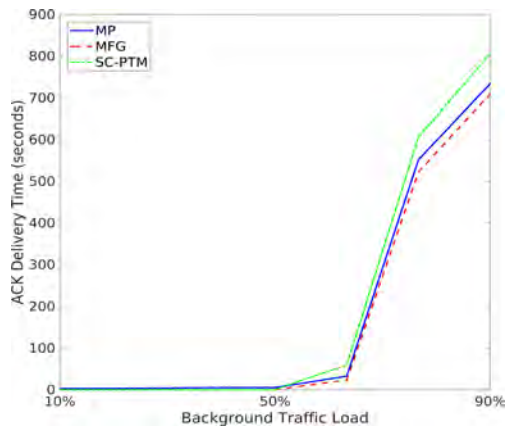


Fig. 18. **ACK Delivery Time with different background traffic loads:** The figure shows the impact of the SC-MCCH in the delivery of the application ACK messages.

energy consumption by keeping the devices waiting for long periods of time.

VI. CONCLUSIONS

In this paper we focused on group-communications in NB-IoT. We discussed a set of enhancements to be introduced in the existing mobile core architecture and new features to be supported by the existing procedures, with particular interest on customer-driven MBMS group formation and content delivery. We also presented two transmission strategies (MFG and MP) aimed at limiting the impact of multicast traffic on unicast transmissions in the downlink and compare against the SC-PTM approach proposed by 3GPP. Our results show that MFG is a valid approach when the network has to handle a large number of multicast sessions as it stops the unicast traffic from overloading the downlink channel. In contrast, the MP approach is more suitable for infrequent multicast sessions, as it allows a multicast content to be delivered quickly without drastically impacting the unicast traffic.

REFERENCES

[1] S. Landstrom, J. Bergstrom, E. Westerberg, and D. Hammarwall, "NB-IoT: A Sustainable Technology for Connecting Billions of Devices," Ericsson, Review, 2016.
 [2] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, and S. He, "Narrowband internet of things: Implementations and applications," *IEEE Internet of Things Journal*, vol. 4, no. 6, Dec 2017.

[3] T. Kim, D. M. Kim, N. Pratas, P. Popovski, and D. K. Sung, "An Enhanced Access Reservation Protocol with a Partial Preamble Transmission Mechanism in NB-IoT Systems," *IEEE Communications Letters*, June 2017.
 [4] S. M. Oh and J. Shin, "An Efficient Small Data Transmission Scheme in the 3GPP NB-IoT System," *IEEE Communications Letters*, March 2017.
 [5] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT system for M2M communication," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016.
 [6] Y. D. Beyene, R. Jantti, K. Ruttik, and S. Iraji, "On the Performance of Narrow-Band Internet of Things (NB-IoT)," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017.
 [7] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements," *IEEE Communications Magazine*, vol. 50, no. 11, November 2012.
 [8] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," 3rd Generation Partnership Project (3GPP), TS 36.331, May 2017.
 [9] J. Schliez and D. Raddino, "Narrowband Internet of Things," Rohde & Schwarz, White paper, 2016.
 [10] Y. P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A Primer on 3GPP Narrowband Internet of Things," *IEEE Communications Magazine*, vol. 55, no. 3, March 2017.
 [11] R. Ratasuk, N. Mangalvedhe, J. Kaikkonen, and M. Robert, "Data Channel Design and Performance for LTE Narrowband IoT," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sept 2016.
 [12] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3rd Generation Partnership Project (3GPP), TS 36.213, September 2016.
 [13] Qualcomm, "Leading the LTE IoT evolution to connect the massive Internet of Things," Qualcomm Technologies Inc., White paper, July 2017.
 [14] 3GPP, "Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (RAN); Stage 2," 3rd Generation Partnership Project (3GPP), TS 25.346, April 2017.
 [15] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over Emerging 5G Networks: Challenges and Perspectives," *IEEE Network*, vol. 31, March 2017.
 [16] A. Urie, A. N. Rudrapatna, C. Raman, and J. M. Hanriot, "Evolved multimedia broadcast multicast service in LTE: An assessment of system performance under realistic radio network engineering conditions," *Bell Labs Technical Journal*, vol. 18, no. 2, September 2013.
 [17] M. Condoluci, G. Araniti, T. Mahmoodi, and M. Dohler, "Enabling the IoT Machine Age With 5G: Machine-Type Multicast Services for Innovative Real-Time Applications," *IEEE Access*, vol. 4, May 2016.
 [18] Cisco, "Multimedia Broadcast and Multicast Service," Cisco, Report, 2016.
 [19] Ericsson, "Massive IoT in the City," Ericsson, White paper, 2016.
 [20] 3GPP, "3GPP Low Power Wide Area Technologies," GSMA Mobile IoT, White paper, October 2016.
 [21] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, "AL-FEC for streaming services over LTE systems," in *2011 The 14th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Oct 2011.