A novel state model for 5G radio access networks

Icaro Leonardo Da Silva, Gunnar Mildh
Wireless Access Networks
Ericsson Research
Stockholm, Sweden
{icaro.leonardo.da.silva, gunnar.mildh}@ericsson.com

Mikko Säily
Nokia Networks

Sofonias Hailu
Aalto University

Espoo, Finland
mikko.saily@nokia.com, sofonias.hailu@aalto.fi

***Abstract*—With the trends towards Internet of Things (IoT) and Machine-Type Communications (MTC) it is expected that the 5th Generation of mobile communications (5G) will have a significant amount of battery powered devices (e.g. sensors, baggage tags, etc.). Therefore, battery efficiency and duration will be essential, especially for those devices in remote locations and/or restricted areas. It would be difficult to predict all the 5G use cases, for example, that may arise from IoT however it is expected that for some of these the tradeoff between efficient power savings modes and low-latency system access, called herein UE sleeping problem might be essential. In order to solve the UE sleeping problem the paper proposes a novel state model for 5G Radio Access Networks (RAN) that relies on a novel state called "connected inactive" where both the UE and the network keep some context information while the UE sleeps. The state is envisioned to be highly configurable in order to address unpredictable use cases possibly with different requirements. It is shown via protocol signaling diagrams that that the proposed solution enables a quick and lightweight transition from inactive to active data transmission.**

Keywords—5G architecture; UE sleeping; State model; Radio Resource Configuration (RRC) protocol; tigh integration; multi-RAT integration;

I. Introduction

Worldwide initiatives in the mobile industry have started in the past few years in order to define the requirements and principles of what is to be the 5th Generation of mobile communications (5G). In all these initiatives (research projects, industry fora, etc.) it has been acknowledged that in order to fulfill the 5G requirements (e.g. in terms of data rates and latency) the Radio Access Network (RAN) should be comprised of at least one novel 5G radio access technology (RAT) tightly integrated with the evolution of Long Term Evolution (LTE), to be widely deployed in the 5G time frame [1], [2]. There is also a consensus in the mobile industry that the overall 5G system should support a diverse range of requirements (sometimes divergent) such as the handling of Massive Machine-Type Communication (M-MTC), low latency (in the order of 1 millisecond) and Ultra-Reliable MTC (U-MTC) and Extreme Mobile Broadband (X-MBB) services with very high peak data rates (up to 20 Gbps) [3][4].

Since the first generations of mobile communications, solutions to optimize the power consumption of mobile devices, called herein User Equipment (UE), have been of key importance. Initially this included solutions using periodic Discontinuous Reception (DRX) cycles and paging to support long standby time for voice centric UEs with reasonable sized batteries. In the 3rd Generation (3G) and in the 4th Generation (4G) of mobile communications, with the advent of smartphones, solutions to quickly switch to DRX in-between data bursts have been standardized [5]. With the trends towards Internet of Things (IoT) and M-MTC it is expected that in 5G there will be even more battery powered UEs (e.g. sensors, baggage tags, etc.). Therefore, battery efficiency and duration will be essential, especially for those devices which accessibility is limited (e.g. remote locations, restricted areas). Efficient ways to enable the mobile devices to switch their receivers off is required, so that they can run on batteries for years. At the same time, the requirement for fast 1st packet transmission from the UEs to the network (or vice versa) is expected to be more stringent in 5G than the previous mobile generations. It is also essential to also reduce the signaling in the network associated with UEs that become active to only send very short data bursts.

Minimizing the latency for 1st packet transmission has the tendency to increase the device battery power consumption. The problem of creating efficient mechanisms that allows the device to minimize its power consumption and, at the same time, be efficiently reachable by the mobile network and quickly start to transmit data with minimum signaling overhead when it requires is called herein "UE sleeping problem".

***Previous contributions* –** Recent contributions have addressed the UE sleeping problem. In [6], [7] and references therein the sleeping problem is either addressed by the optimization of existing DRX parameters in LTE or by enhancing the mechanism via new procedures. In [6], a method to reduce the power consumption during DRX "on periods" for MTC devices in LTE is proposed. The method relies on avoiding Physical Downlink Control Channel (PDCCH) decoding, a computationally intensive process that requires substantial processing time that is not useful when the device is not paged. This is achieved thanks to a Quick Sleeping Indication (QSI) independently from PDCCH that allows the device to sleep earlier based on the assumption that MTC devices using LTE are unlikely to be paged every time it wakes up. In [7], the authors have studied the tradeoff between latency and power savings for DRX configured UEs in Radio Resource Control (RRC) CONNECTED state and proposed optimizations of the DRX parameters. It has been shown that DRX short cycles are very effective in reducing latency for active traffic, while shorter inactivity timers are desirable for background traffic to enhance power saving. The authors have also proposed a mechanism to switch between DRX configurations based on different traffic conditions at the UE, using UE assistance (adopted by 3GPP in Release 11). Despite its proven benefit, the solution proposed in [7] is applicable for UEs in RRC CONNECTED state where mobility is fully network based (i.e. requiring constant signaling between the UEs and the network). In the 5G time frame billions of devices are expected so that it can be very challenging (and inefficient) to keep them all UEs in RRC CONNECTED with full network controlled mobility support otherwise network signaling will explode.

Our contribution – This paper addresses the "UE sleeping problem" by proposing a novel state model for the 5G RAN (more specifically, to the RRC protocol). The proposed state model comprises a novel state called "RRC CONNECTED INACTIVE", in addition to the existing ones (i.e. "RRC IDLE" and "RRC CONNECTED"). The paper presents the characteristics of the novel state and how it improves the existing state model in LTE by [8]:

- Reducing the delay that takes for the UE to access the system and start the data transmission after inactivity periods.
- Reducing the signaling overhead in the radio and the Core Network (CN)/RAN interfaces for UEs coming from inactivity periods that want to transmit small amounts of data (e.g. MTC devices) or short data bursts (e.g. some types of smartphone traffic) after periods where the UE receiver was off.
- Reducing the signaling for moving UEs by enabling an efficient UE-based mobility mechanism during the inactivity periods.
- Enabling a higher level of configurability in order to address a wide range of use cases and services that may have divergent requirements (e.g. in terms of accessability and power consumptions), expected in the 2020 time frame.
- Enabling configurable multi-RAT features (such as multi-RAT camping) where the evolution of LTE is tightly integrated to the 5G RAN.

These benefits are achieved partially by exploring the principle of "not discarding previously exchanged information" for inactive UEs. In other words, UEs moving to the "RRC Connected Inactive" state will keep parts of the RAN context. This is valid at least for the *semi-static information* i.e. still valid after inactivity periods such as Access Stratum (AS) security context, UE capability information, etc. In addition to storing the RAN context the UE is allowed to move around within a pre-configured area without notifying the network. This avoids the need in the network to handle the mobility of these devices. A similar principle based on storing the RAN context has been proposed in 3GPP to support highly efficient handling of frequent and infrequent small data transmissions with minimized overhead for system signaling but for LTE RRC Idle UEs [15].

Paper organization – The paper is organized as follows. Section II describes in more details the "UE sleeping problem". Lessons learnt from LTE networks are presented in order to motivate the proposed solutions. In Section III, the principles of the novel RRC_CONNECTED_INACTIVE state are presented. The impact of the novel state to the 5G RAN tightly integrated to the evolution of LTE is also analyzed. In Section IV, the overall state model for the 5G RAN is presented where some of the benefits highlighted are shown via protocol signaling diagrams.

II. UE Sleeping Problem

In order to enable the UE to save battery it is very important that the UE can switch off its receiver (and transmitter) when there is no data to be transmitted or expected for that UE. This makes possible to achieve significantly longer standby time in the UE compared to "talk" (or active) time. At the same time the UE should be reachable by the network (e.g. via paging) and, if it wants to transmit data, it should quickly be able to access the system. This tradeoff between power savings and quick system access is called herein "UE sleeping problem".

A. Discontinous Reception (DRX)

In current systems the UE sleeping problem can be controlled via DRX optimizations [5]. A DRX cycle consists of "on periods" during which the UE monitors downlink channels and physical signals (so that it can be reached by the network via paging among other actions) and "sleeping periods" when the UE can switch its receivers off. When the UE is utilizing DRX it probably also use Discontinuous Transmissions (DTX) however this is not specified in the standards. The tradeoff between power savings and access latency can be controlled by optimizing the settings for the DRX cycles. In the one hand, DRX allows the UE to sleep and, consequently save battery when receivers are off.

B. DRX applied to the existing states

Currently, LTE supports two different states in which DRX can be configured: a CN state where the UE location is not known in the RAN and a RAN connected state where the UE has an ongoing RAN connection [8]. Some of the characteristics in these two states are the following:

- In the CN state, called Evolved Packet System (EPS) Connection Management (ECM) IDLE / RRC IDLE, only CN context is stored when the UE is sleeping. UE and network discards the RAN context information when moving to this state. The UE is known at the network only on tracking area level and may move within the cells belonging to the tracking area without informing the network. The UE would, nevertheless, camp in the best cell via cell reselection procedure based on the configuration provided by the network. In this state DRX is used to improve battery performance so the UE is not mandated to constantly monitor paging channels and transmit location updates if highly moving. Therefore, DRX mainly impacts the paging performance where longer DRX cycles (for UE power savings) are traded off with the reachability of the UE by the network (that needs to be on to monitor paging channels).

- In the RAN connected state (ECM CONNECTED / RRC CONNECTED) the UE is known on a cell level and mobility is fully network controlled (via handovers). RAN context is present. In this state DRX is used for micro-sleeping periods between bursts of data of better resource utilization. For instance, during a web session short DRX cycles can allow for faster response when data transfer is resumed [7] instead of wasting resources to continuously monitor downlink channels (e.g. PDCCH) while the user is reading a downloaded web page. On the other hand, due to the uncertainties of the traffic demands, DRX can lead to delays when the network wants to reach the UE (e.g. if there is incoming traffic) [10]. In order to reach a good compromise current cellular systems such as LTE supports two configurable DRX cycle, a long and a short with fast switching in between. For Uplink (UL) traffic the added delay from DRX is less of a problem since the UE can in principles request resources as soon as UL data arrives (assuming there are available scheduling request resources).

C. *Lessons learnt from LTE networks*

It can be said that out of the two states explained earlier RRC IDLE is used as the primary sleeping state since it has been observed in LTE networks that inactivity timers are typically configured to be quite short (between 10-60 seconds). As a consequence, a very high amount of transitions from RRC IDLE to RRC CONNECTED are observed in existing networks. One problem is that this state transition is quite costly in terms of signaling especially considering that the majority of the RRC connections transfer less than 1 Kbyte of data to then move back to RRC IDLE [9][10].

The signaling diagram of Fig. 1 shows this state transition, required for the UE to transfer/receive user plane. This overhead may also introduce significant delays. In the best case, without taking into account the processing delays at the UE, network side and signaling within the CN (e.g. towards the Serving Gateway), this overhead roughly introduces a delay, measured in terms of Random Access (RA) and Round-Trip-Times (RTT) of:

Transition time >RA delay + 4 x RTT (radio) + RTT (S1).

In LTE there is a requirement that this delay should be lower than 100 ms and, in LTE-A, lower than 50 ms, so that even lower values should be expected in 5G at least for some services [11]. This scenario is also expected to exist in 5G networks with the trends towards IoT [16] where it is expected a large number of devices, each generating a small amount of data. As it can be observed, most of the radio signaling comes from random access and RRC Connection Setup related procedures while the inter-node signaling, between the RAN and the CN, comes from the whole setup of the S1 connection, UE capability information exchange, etc.

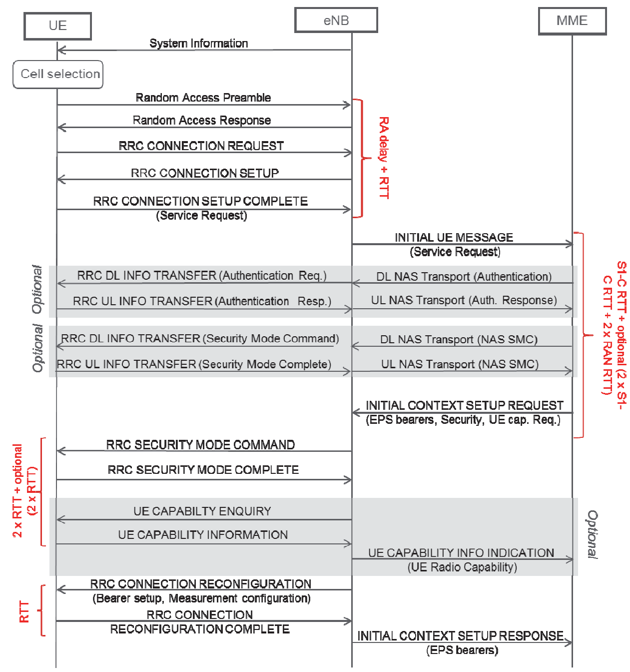

Fig. 1. Signaling for the RRC IDLE to RRC CONNECTED transition, which is the most typical way to move from inactivity to start transmitting data.

III. PRINCIPLES OF THE NOVEL CONNECTED INACTIVE STATE

As described in Section II, initiating data transmission from RRC IDLE in LTE involves significantly more signaling (and consequently delays) compared to data transmission from RRC CONNECTED with DRX (LTE requires the DRX transition in RRC CONNECTED to be lower than 10 ms [11]). On the other hand, it is challenging and not efficient to always keep all UEs in RRC CONNECTED state where procedures are optimized for data transmission such as network-controlled mobility, especially assuming that there will be billions of devices in 5G networks.

In order to address the "UE sleeping problem" for 5G devices it is proposed a novel "RRC CONNECTED INACTIVE" state designed to be used as the primary sleep state for the 5G access. This new state will have the following characteristics:

- **Widely configurable DRX** cycles (from milliseconds to hours) to support a wide diversity of services with different requirements in terms of power consumption and accessability delays.
- **UE controlled mobility and RAN-based paging** with optimized state transitions for the case where the UE is semi-static i.e. the UE remains in the same location

after inactivity timer expires. The concept of camping for IDLE UEs is extended to the RRC CONNECTED INACTIVE UEs.

- **Configurable multi-RAT procedures** (such as multi-RAT camping) where the evolution of LTE is tightly integrated to the 5G RAN.
- **Highly configurable procedures** that may possibly take into account known characteristics at the RAN level such as mobility pattern and traffic characteristics for the different services and performance requirements in terms of delay accessability.

A. Widely configurable DRX

The wide diversity of 5G use cases will also lead to devices with very different traffic patterns and battery requirements. Since the RRC CONNECTED INACTIVE is envisioned to be used as the primary sleeping state it is essential to enable a widely configurable DRX cycles in order to comprise the different cases. Some devices may need to sleep for hours and minutes, while others would need to wake up only once a day but still benefit from the fact the UE context is stored e.g. for network-initiated contact.

B. UE-controlled mobility and RAN-based paging

It is envisioned that in the novel “RRC CONNECTED INACTIVE” the UE is reached by the network via paging so that the UE can be configured to monitor the paging channel(s). Since this is a RAN state, in order to be reached it is also envisioned that the UE is known at the RAN within what so-called Tracking RAN Area (TRA) where mobility signaling is avoided when the UE moves within a configured TRA or TRA list (possibly similar to an LTE UE in RRC IDLE). This would optimize the amount of signaling that could be from moving M-MTC UEs.

C. Configurable multi-RAT procedures

It has been acknowledged that in order to benefit from its widely deployed coverage in the 2020 time frame LTE should be tightly integrated to the 5G RAN [2][3]. A solution relying on a common PDCP and RRC frameworks has been recently proposed [1], inspired in previous research in the area of multi-RAT integration [12]. This will enable the dynamic usage of all available resources (free resources in other access). If this takes long time, those free resources might no longer be there. These scenarios also motivate a common CN connection (e.g. an evolved S1) for both accesses in the case of dual-radio UEs.

A common control plane framework where the evolution of LTE is part of the 5G access demands a common state handling for dual radio UEs. Otherwise, any toggling between the novel 5G RAT and LTE coverage due to bad coverage reasons would lead to signaling to update the LTE state.

In “RRC CONNECTED INACTIVE” the UE can be configured to either camp on LTE or on the novel 5G RAT (and monitor the respective paging channels accordingly). The main driver for the RRC_ CONNECTED_INACTIVE state in LTE for the dual-radio UEs is the tight integration. However, one could argue that similar solutions also provide benefits for only-LTE UEs. Solutions following a similar principle have been recently proposed for LTE [15].

D. Highly configurable procedures

The network can explore the fact that the UE context is kept at the RAN and perform some optimizations for the procedures in the RRC CONNECTED INACTIVE state e.g. based on known characteristics about the services and/or mobility patterns. For instance, by knowing that a given UE is static, the network could maintain the mobility anchor point in a single node so that paging only occurs in a limited area. For certain services, where access is only UE-initiated, paging may not even be configured.

Depending on how predictable the traffic is and the service requirements in terms of access latency the network can configure certain UEs to access the system with some prioritized mechanisms (e.g. RACH configuration dedicated for groups of UEs) since resource ownership is better handled at the RAN. Another possibility could be to configure some sort of multi-RAT camping for dual-radio UEs with high accessability requirements. In that sense the UE could either monitor paging channels on LTE and the novel 5G RAT to improve reachability. The impact on UE battery performance should be carefully considered in that case.

IV. State model for The 5G Access

A novel state model is proposed for the 5G Access enabling an efficient UE sleeping, a fast and lightweight transition from sleeping to active states and joint access optimizations. The RAN part of that model is shown in Fig. 2.

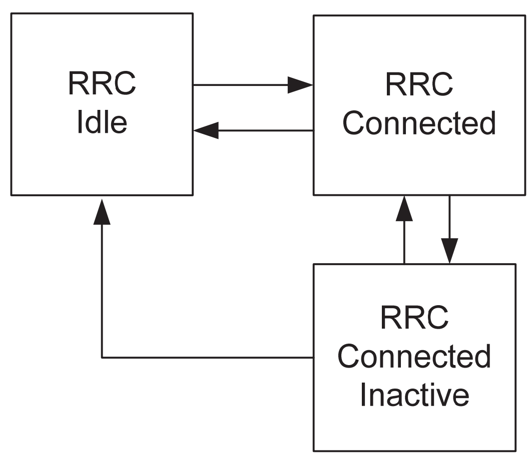

Fig. 2. State model for the 5G architecture (only RAN states)

The model consists of three states: RRC_IDLE, RRC_CONNECTED and RRC_CONNECTED_INACTIVE. In the novel model the state transitions from RRC_IDLE → RRC_CONNECTED are expected to occur mainly during the first initial access (e.g. when the UE attaches to the network) or as a fallback case (e.g. when the devices and/or network cannot use the previously stored RAN context). As a consequence, this transition is not expected to occur as often as in LTE.

On the other hand, transitions from RRC CONNECTED INACTIVE → RRC CONNECTED are expected to occur quite often and should be optimized as a lightweight and fast transition. Some aspects of this transition are described in the following. Details of this procedure are described in the following.

A. Fast and lightweight transition from Inactive to Connected

In the proposed state model the RRC CONNECTED INACTIVE is the primary sleeping state i.e. optimized for data inactivity. When the UE moves to that state (e.g. via an inactivity timer or via explicit network signaling) both the UE and the network will keep RAN context information related to both accesses (i.e. the novel 5G RAT and LTE) that has been obtained during the first RRC connection setup e.g. when the UE attaches to the network and/or moves from RRC_IDLE → RRC_CONNECTED. One example of context information is the UE capabilities, both related to the 5G RAT and LTE, and security context that could be considered as semi-static.

It is envisioned that this state transition is handled by a procedure inspired in what is being defined in [10], called RRC Resume. Herein this involves the resumption of Signaling Radio Bearers (SRB) and Data Radio Bearers (DRB). The connection re-activation succeeds only if the accessed target node (LTE or the novel 5G RAT) can find the UE RRC context and the mobility anchor for the CN/RAN interface. For this reason, some sort of UE RRC Context ID should be included within a *RRC Connection Re-activation Request* that is an SRB0 message**.** This message could be integrity protected to protect the network for denial of service attacks or false requests.

The procedure is triggered by the UE either in response to a paging, when the UE has uplink data in buffer or when it needs to send TRA updates. The UE triggers an RRC connection re-activation procedure which should be defined in both new 5G RAT and LTE's RRC specifications. Upon receiving *RRC Connection Re-activation Request*, the network retrieves the UE RRC Context (including the security re-activation information) based on the UE RRC Context ID, performs the necessary mobility actions and responds with *RRC Connection Re-activation* to reconfigure SRBs and DRBs. Upon the reception the UE performs the following actions:

- Re-establish PDCP and RLC for SRBs and DRBs;
- Perform radio resource configuration;
- Perform measurement related actions according to the measurement configuration;
- Resume SRBs and DRBs.

A simplified delay calculation for the proposed procedure (as shown in Fig. 3) is the following:

Transition time >RA delay + 1.5 x RTT (radio).

As it can be noticed the RTT is reduced from 4 to 1.5 RTT while there is no need to setup the CN/RAN connection, assuming negligible delays to fetch the RAN context in the network side, which is reasonable considering that most of the transitions may occur in the same area. When it comes to absolute values a potential difference compared to LTE may exist since what is called RA delay comprise the delays from procedures (such as frequency and time synchornozation) and random accesses that may be different in the novel 5G RAT.

The state transition based on RRC signaling should be seen as a default option that needs to be support however, further optimizations may also exist e.g. some MAC-based signaling in the case of semi-static devices.

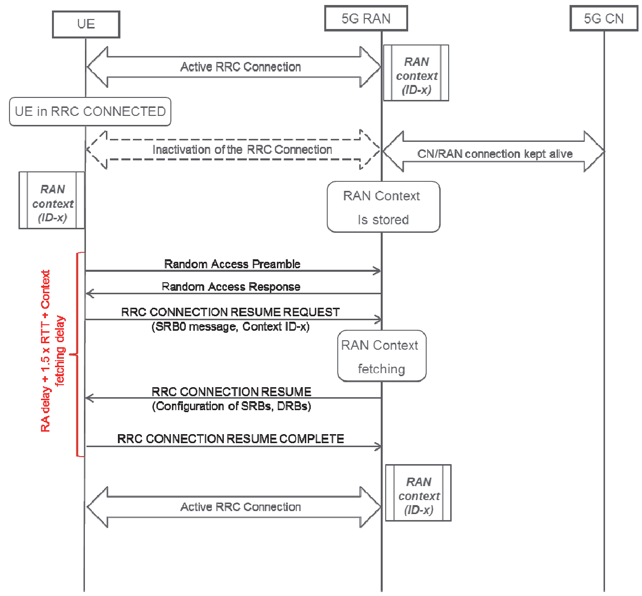

Fig. 3. Signaling for the RRC CONNECTED INACTIVE to RRC CONNECTED transition for the novel state model.

B. Transparent inactivity to the CN

It is also worth mentioning that in the propose state model the CN/RAN context is also kept i.e. the transits from and to RRC CONNECTED and RRC CONNECTED INACTIVE are transparent to the CN. Therefore, incoming packets from the CN may be forwarded to the latest mobility anchor point at the RAN so that transitions from RRC CONNECTED INACTIVE to RRC CONNECTED do not need to involve CN signaling, which may lead to longer delays and network signaling.

V. CONCLUSION

The sleeping problem for 5G devices has been addressed by the proposal of a new state model enabling an efficient UE sleeping, a fast and lightweight transition from sleeping to active states and a fast establishment of dual connectivity in order to fulfill 5G reliability requirements. This new state model contains what we call an RRC CONNECTED INACTIVE that explores the principle of "not discarding previously exchanged information" for sleeping UEs i.e. UEs in RRC CONNECTED INACTIVE state will keep parts of the RAN context as recently proposed in 3GPP for LTE [15]. Another novel aspect is that the proposed state model has been designed to efficiently support the tight integration of LTE and

NX [1] optimizing the procedures needed to support dual connectivity.

We can also conclude that the assumption of a common CN and a common S1 connection for both accesses (as discussed in Section III) reduces the amount of network signaling needed between the CN and the RAN for either the establishment of dual connectivity or the transition from the sleeping state to any of the active states. A consequence of this overhead reduction is the reduction of the delay from the sleep state to any of the active states. This might be quite beneficial in scenarios where the NX link can be quickly lost (since NX operates in higher frequency bands and rely massively on narrow beamforming [13]), in the case of a fast dual connectivity or in scenarios where the UE needs to quickly establish a data connection to transmit a small amount of data, as in the case of many envisioned MTC use cases for 5G.

ACKNOWLEDGMENT

This work has been performed in the framework of the H2020 project METIS-II co-funded by the EU. The views expressed are those of the authors and do not necessarily represent the project. The consortium is not liable for any use that may be made of any of the information contained therein.

REFERENCES

[1] I. Da Silva, G. Mildh at al., "Tight integration of new 5G air interface and LTE to fulfill 5G requirements," Vehicular Technology Conference (VTC Spring), IEEE 81st, pp. 1-5, Glasgow, May 2015.

[2] P. Marsh, O. Queseth et al., "5G radio access network design – a brief overview on the 5G-PPP Project METIS-II", European Communicatoin Networking Conference (EuCNC), Paris, July 2015.

[3] NGMN Alliance, "5G White Paper", February 2015, available at https://www.ngmn.org/fileadmin/ngmn/content/images/news/ngmn_news

[4] MET14-D43 ICT-317669-METIS/D8.4 "METIS final project report " April 2015

[5] Evolved universal terrestrial radio access (E-UTRA); medium access control (MAC); protocol specification," 3GPP TS 36.321, Rel. 10, V.10.2.0, Jun. 2011

[6] N. M. Balasubramanya, L. Lampe. G. Vos, S. Bennett, "Introducing quick sleeping using broadcast channel for 3GPP LTE MTC", Globecom workshops – Telecommunications Standards – from research to standards, December 2014.

[7] A. T. Koc, S. C. Jha, R. Vannithamby, M. Torlak, "Device Power Saving and Latency Optimization in LTE-A Networks Through DRX Configuration", IEEE Transactions on Wireless Communications, Vol. 13, No 5, May 2014.

[8] Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," 3GPP TS 36.331, Rel. 10, V.10.3.0, Sept. 2011.

[9] "LTE RAN enhancements for diverse data applications," 3GPP TR 36.822 V0.2.0, Nov. 2011.

[10] "Study on non-MTC mobile data applications impacts (Release 12)," 3GPP TR 22.801, Rel. 12, V.12.0.0, Dec. 2011.

[11] "Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)", 3GPP TS 36.913, Rel. 12, Sep. 2014.

[12] M. Johnsson, J. Sachs, T. Rinta-aho, T.Jokikyyny, "Ambient networks – a framework for multi-access control in heterogeneous networks", in proc. IEEE Vehicular Technology Conference (VTC), Montreal, Canada, Sep 2006.

[13] Ericsson Research blog, "Massive beamforming in 5G radio access", 2015,http://www.ericsson.com/research-blog/5g/massive-beamforming-in-5g-radio-access/

[14] Ericsson Research blog, "Release 14 – the start of 5G standardization", 2015, http://www.ericsson.com/research-blog/lte/release-14-the-start-of-5g-standardization

[15] "S2-152207 Signaling reduction for UE states transitions", 3GPP SA2 WG, Dubrovinik, Croatia, June 2015.

[16] "Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things; (Release 13)", 3GPP TR 45.820 v1.3.1 (20115-06).