mIoT Connectivity Solutions for Enhanced 5G Systems

M. Condoluci*, R. Trivisonno†, T. Mahmoodi*, and X. An†

*Centre for Telecommunications Research, Department of Informatics, King's College London, UK

e-mail: {massimo.condoluci, toktam.mahmoodi}@kcl.ac.uk

†Huawei Technologies, Munich, Germany

e-mail: {riccardo.trivisonno, xueli.an}@huawei.com

Abstract—Within the ongoing activities devoted to the definition of 5G networks, massive Internet of Things (mIoT) is regarded as a compelling use case, both for its relevance from business perspective, and for the technical challenges it poses to network design. With their envisaged massive deployment of devices requiring sporadic connectivity and small data transmission, yet QoS constrained, mIoT services will require ad-hoc end-to-end (E2E) solutions, i.e., featuring access and core network enhanced Control and User planes (CP/UP) mechanisms. This paper presents and evaluates a novel connectivity solution to manage massive number of devices. The paper presents an analytical model developed to evaluate the performance of the proposed solution. Quantitative results derived from the model demonstrate the effectiveness of the solution proposed in this paper, compared to 4G systems, and its ability to reduce CP signaling and optimize UP resource utilization for massive device deployment.

Index Terms—5G Systems, mIoT, Core network.

I. Introduction

Besides achieving a challenging set of key performance indicators (KPIs), the definition of 5G networks has been driven by the attempt to conceive a communication system allowing the integration of vertical industries, including massive Internet of Things (mIoT) as a prominent example [1]. The attention devoted to mIoT has been constantly increasing over the past years, as market reports and data traffic forecasts predicted a proliferation of a wide variety of applications and services which, together with the massive deployment of e.g. sensors, meters, actuators, wearables and connected appliances, will lead to a 75% increase of wireless mobile connectivity request by the end of the decade. However, despite the high industrial interest, due to the enormous required effort, standardization bodies are set to finalize an early 5G release in 2018 which, on one hand, will provide the key innovation elements necessary for verticals integration [2] (namely architecture modularization, network slicing and service based interfaces) [3], [4], [5], [6], [7] but, on the other hand, will specify only the support of enhanced Mobile Broadband (eMBB) services. The support of mIoT services, hence, will be included in enhanced 5G systems, which will be defined in 3GPP Release 16 specifications.

The proper support of mIoT in 5G systems strongly depends on the capability of taking into consideration the needs of IoT (e.g., very long battery duration) as well as its features (e.g., small packets with periodicity up to a few hours, high device density) into the design of the network. This dictates for ad-hoc solutions in both radio access and core networks (RAN and CN, respectively). The RAN has been attracting the attention of the research community, in particular to improve the performance of the random access (RA) procedure used by the user equipments (UEs) to trigger a connection request [8]. Several solutions have been proposed to increase the number of UEs supported in the RA (e.g., [9]) or to limit the congestion in the RAN via access class barring (ACB) or to reduce the control signaling sent over the air [10] to save devices' battery and improve spectrum utilization. From a 5G perspective, envisaged mIoT traffic models and deployment scenarios do not represent a serious challenge on the RAN as solutions proposed by 3GPP (a.k.a. 4G CIoT [11]) appear to be suitable in the RAN as massive deployment of devices goes in pair with sporadic data transmission.

Less attention has been dedicated to the CN, relying on the connectivity mechanisms of 4G systems. Nevertheless, it is worth mentioning that the current connectivity model of 4G systems is based on the idea that the CN manages the connectivity on a per-UE basis, with consequent load issues in terms of number of connections to be simultaneously managed in the CN due to the huge density of IoT devices. It is worth noticing that the device density issue is accentuated in the CN compared to the RAN, as thousands of RAN nodes are connected to the same CN.

This paper proposes a novel end-to-end (E2E) connectivity solution for mIoT support in enhanced 5G systems, designed in compliance to the reference architecture of 3GPP Release 15 [5]. The presented solution is based on the idea of an Aggregate Core Network Bearer (ACNB), which is introduced to carry the traffic from/to multiple UEs with similar communication requirements. The solution covers both control and user planes (CP and UP, respectively), and aims to reduce the CP signaling in the CN while ensuring an efficient utilization of UP resources without affecting the capability of the network to satisfy the diverse QoS that different device categories might require. An analytical model is proposed to evaluate the performance of the presented connectivity solution on both CP and UP while taking into consideration the impact of RAN on the CP/UP traffic generated towards the CN, where results demonstrate two digit gain compared to 4G. The results also
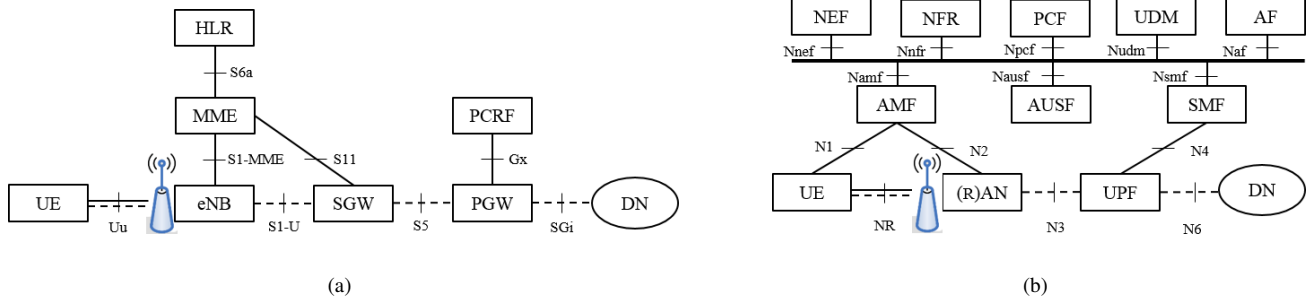
Fig. 1. LTE/EPC reference architecture (a) and 5G System Service Based Reference Architecture (b).

shows that, according to the considered device deployment scenarios and traffic models, massive connectivity as per 5G requirements may turn to be critical at CN rather than at the RAN, thus further motivating the focus on this study on the CN side.

The paper is structured as follows. After a review of 4G and early 5G connectivity mechanisms included in Section II, Section III provides a detailed description of the proposed connectivity solution for mIoT in enhanced 5G systems. Section IV describes the analytical model defined to assess the solution performance, and Section V reports the results of the quantitative performance evaluation. Final remarks are given in Section VI.

II. CONNECTIVITY MODELS IN 4G AND EARLY 5G SYSTEMS

A. EPS Bearer in 4G Systems

The 4G LTE/EPC Evolved Packet System (EPS), depicted in Fig. 1(a), relies on a connectivity model based on the "*EPS bearer*" and the "*Always-ON*" concepts [12]. The EPS is designed to provide IP connectivity between a UE and a PLMN external Data Network (DN). The EPS bearer is the minimum level of granularity at which QoS, mobility and security are provided within EPS: when a UE attaches to a DN, after authentication, it is allocated an IP address and an EPS bearer is established. The EPS bearer remains established throughout the lifetime of the DN connection to provide the UE with Always-ON IP connectivity. Fig. 2(a) illustrates how IP connectivity is established in 4G systems.

The bearer established at the attachment is the default bearer that provides default QoS determined upon UE subscriber data. Any additional EPS bearer is referred to as a dedicated bearer, established when either the UE or the network issue a Service Request, for which a dedicated QoS treatment needs to be provided. The EPS bearers are established by LTE/EPC CP, and they are E2E concepts: they provide connectivity from UE to the PLMN external DN, hence traversing UP elements of RAN and CN. In particular, the EPS bearer is composed by the concatenation of the radio bearer (between the UE and the access node), the S1 bearer (between the access node and the Serving Gateway, SGW) and S5/S8 bearer (between SGW and Packet Data Gateway, PGW).

B. PDU Session and per Flow QoS in Early 5G Systems

Even if with some differences due to the new 3GPP release 15 architecture depicted in Fig. 1(b), the early 5G system (5GS) defines a connectivity model not dissimilar from 4G, this justified by early 5G supporting uniquely the eMBB service. The 5GS supports connectivity service via PDU sessions [5] that are established upon request from the UE and provide exchange of PDUs between a UE and a data network (DN). Three types of PDU sessions are defined: IP (v4/v6), Ethernet and Unstructured. After successful registration to the 5GS (handled by the Access and Mobility Management Function, AMF), PDU sessions can be established, modified and released via non-access stratum (NAS) signaling exchanged over N1 interface between the UE and the Session Management Function (SMF). For each active PDU session a UE-CN UP connection is established, comprising a data radio bearer between the UE and the access node (AN) and an N3 tunnel AN-UPF. Unlike 4G, a registered UE may or may not have an active PDU session.

III. mIoT CONNECTIVITY SOLUTION FOR ENHANCED 5G SYSTEMS

A. Design Rationale

From a high level but not superficial comparison between 4G EPS and 5GS, it clearly emerges that both systems rely on connection-oriented connectivity models and foresee at least a tunnel per UE between eNB/AN and SGW/UPF. If, on one hand, this design choice appears to be reasonable for eMBB service, on the other hand envisaged device deployment scenarios and traffic models expected for mIoT service may make the validity of the models questionable.

An educated assessment of such models needs to consider mIoT deployment scenarios usually envisage a device density between 10^3 and 10^5 UE/km^2, leading up to 10^6 devices per cell, and traffic patterns (either synchronous or asynchronous) predominantly device triggered, with different access distributions in time (e.g., uniform, beta-distributed [10]) over periods ranging from tens of seconds to tens of hours. The UP traffic generated is hence expected mostly on the uplink (UL), each device activation requiring the transmission of data ranging between 10^2 and 10^4 bits. This would require networks
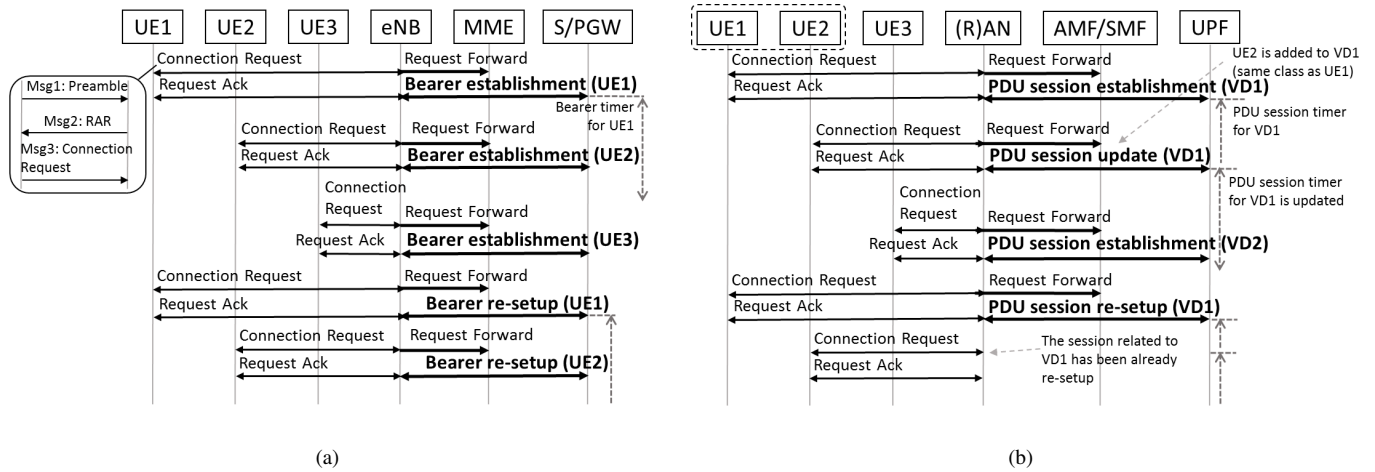
Fig. 2. 4G Connectivity (4G-CM) (a) and enhanced 5G Connectivity for mIoT (mIoT-5G-CM) (b) establishment procedures.

to handle number of connections 3-4 orders of magnitude higher than Mobile Broadband (MBB) use cases, while each connected device will lead to a data traffic 5 to 8 orders of magnitude lower than 4G smartphones. Additionally, it shall be noticed that bearers/PDU Sessions establishment requires time, computation and storage resources for CP (eNB/AN and MME/SMF) and UP (eNB/AN and SGW/UPF) entities [13], [11]. 4G UP network elements SGW/PGW (and, likely, early 5G UPFs) can simultaneously handle only a target maximum number of bearers ($\sim 10^6$-10^7) and their complexity and cost increases accordingly. Considering 4G, the massive CP load together with the huge number of UP connections determined by mIoT scenarios would require either enhanced (and more expensive) EPC network elements, or network re-engineering, re-planning and re-deployment. Even with an approximate quantitative evaluation, thousand of LTE cells can lead to up to one billion of EPS attached mIoT devices and assuming best practice 4G network topologies, this goes clearly beyond EPC network element capabilities. In addition, even assuming enhanced CP and UP capable of handling the required number of bearers, the approach would lead to an extremely inefficient resource utilization at EPC as IP connectivity would be established for devices actually idle most of the time. The Always-ON concept presumes the UE, while attached, will require frequent data exchange. This justifies the resources allocated to support the default bearer. Such assumption clearly does not apply to mIoT, where small and infrequent data transmission appears to be the dominant model which can be managed in the RAN through the 4G CIoT solutions [11].

B. Solution Description

The design of a connectivity mechanism to support mIoT in enhanced 5GS starts from the need for a connection-oriented model satisfying QoS requirements of mIoT services and, at the same time serving, efficiently, massive number of connections both from CP and UP perspective. As highlighted in the review of 4G and early 5G systems included in Section I

and III.A, 4G CIoT enhancements [11] already allow to tackle the challenges on the access posed by the described traffic models. Nevertheless, to achieve the goals described above, some new concepts affecting the CN are introduced.

First, the *device class* is defined, which identifies the devices communication requirements (e.g., QoS profile, reliability, availability, supported DN, DN specific requirements). Devices belonging to the same class are characterized by homogeneous communication requirements and hence, their data transmission requires a homogeneous treatment by the network. Second, the *Virtual Device (VD)* concept is introduced: a VD is a logical entity at CN side which corresponds to the aggregation of a number of UEs camped on the same cell and belonging to the same device class. The VD concept allows a multitude of UEs belonging to the same class to be handled at CN as a "single device", whose behavior is regulated by a single state machine. For a given device class, a maximum number of UEs can be associated to a VD, where this number can be tuned by the network. A VD inherits the same device class of the devices it gathers.

Finally, the concept of *Aggregate Core Network Bearer (ACNB)* is introduced, transporting UP data from the (R)AN function through the CN UPFs until the external DN. An ACNB is associated to each VD, and it transports data from the (R)AN/DN Gateway to the DN Gateway/(R)AN generated by/from all UEs composing the VD. In other words, the ACNB makes a single simplified UP for a plurality of UEs for which data packets require the same treatment in terms of QoS, reliability, availability, etc. For a given device class, each ACNB is associated to an activity timer, restarted at each data transmission. Upon the expiration of the activity timer, the ACNB is released and the VD is de-allocated.

Fig. 2(b) illustrates the introduction of the proposed connectivity model within 5G System. The diagram shows how the ACNB is established when a first device requires connectivity via a given AN, and how the same ACNB is used to tunnel

UP data for other mIoT devices of the same class connecting via the same AN.

IV. Modeling Connectivity for Enhanced 5G Systems

For the evaluation of 4G and proposed 5G connectivity mechanisms, an analytical model has been developed. To obtain significant results, the RAN segment needs to be modeled in a very detailed way, as the RAN performance impacts the CP/UP traffic sent towards the CN. Table I lists the notations used by the analytical model.

The set $\mathcal{P} = \{1, 2, \ldots, P\}$ indicates the UEs active during D. Different device classes are considered, each with a different QoS and thus bearer/ACNB configuration. Each UE, $p \in \mathcal{P}$, is associated to a class $c = \{1, 2, \ldots, C\}$. The UEs associated to class c are denoted with $\mathcal{P}_c \subseteq \mathcal{P}$, where $\mathcal{P}_x \cap \mathcal{P}_y = \emptyset$, for $x \neq y$ and $\cup_{c=1}^{C} \mathcal{P}_c = \mathcal{P}$.

Each UE, p, is associated to a VD with class c. The parameter $c_v \in \{1, 2, \ldots, C\}$ is the class of the v-th VD, where $v = \{1, 2, \ldots, V\}$ and V denotes the number of VDs activated by the CN. The set $\mathcal{K}_v$ indicates the subset of UEs associated to v-th VD, where $\mathcal{K}_v \subseteq \mathcal{P}_{c_v}$, i.e., multiple VDs can be activated to serve the same class c. A UE can be associated to only one VD, i.e., $\mathcal{K}_x \cap \mathcal{K}_y = \emptyset$ with $x \neq y$. Of course, $\cup_{v=1}^{V} \mathcal{K}_v = \mathcal{P}$.

Bearer/ACNB could have different lifetimes T_c, which depends on the class c. If there is no UP traffic for an interval equal to T_{c_v} for the bearer/ACNB associated to the v-th VD, a re-setup will be required in case of a connection request from a UE, p, belonging to VD, v.

The parameter U_{c_v} indicates the maximum number of UEs that can be associated to a single ACNB, which depends on the device class c_v. In the 4G connectivity model, it is $U_{c_v} = 1$ for any device class.

The model evaluates the CP signaling generated by a RAN node towards the CN taking into consideration collisions in the access, transmission failures, limitation of wireless resources, and load of the RAN. The connectivity request is triggered by a UE by means of a random access (RA) [8][1] that consists of: *Msg 1*, i.e., the transmission by the UE of a RA resource (an orthogonal preamble in 4G) randomly selected among R RA resources; *Msg 2*, i.e., a feedback sent by the RAN node related to *Msg 1* (RA response, RAR, in 4G); *Msg 3*, i.e., the transmission by the UE of a connection request; *Msg4*, i.e., an acknowledgment by the RAN indicating that the connection has been successfully established. The periodicity of RA opportunities is T^{RA}. For the sake of simplicity, the evaluation focuses on the traffic related to bearer/ACNB for each RA opportunity $i = 1, 2, \ldots, I$, where $I = \lceil D/T^{RA} \rceil$.

Let's focus on a generic i-th RA slot. The UEs performing the RA procedure in this slot are denoted with $\mathcal{A}_i \subseteq \mathcal{P}$. The maximum number of attempts for the RA is N, $n_p = 1, 2, \ldots, N$ denotes the counter of RA attempts for p-th UE

[1]For the sake of simplicity, only one RAN node is considered but the model can be easily extended to multiple RAN nodes attached to the same CN.

Notation	Definition	Value
$\mathcal{P}$	Set of UEs	-
$\mathcal{P}_c$	Set of UEs belonging to class c	-
$\mathcal{K}_v$	Set of UEs associated to VD v	-
$\mathcal{A}_i$	UEs performing a RA in the i-th RA slot	-
$\mathcal{B}_i$	UEs receiving the *Msg 2* in the i-th RA slot	-
$\mathcal{C}_i$	UEs transmitting the *Msg 3* in the i-th RA slot	-
C	Number of bearer/ACNB classes	-
V	Number of activated VDs	-
D	Duration of the interval of interest	$1h$
T_c	Lifetime of a bearer/ACNB with class c	$2min$
U_c	Maximum UE number for a VD with class c	100
T^{RA}	Periodicity of RA opportunities	$5ms$ [10]
I	Number of time slots in the time interval D	-
N	Maximum number of RA attempts	10 [10]
R	Number of resources for the RA	54 [10]
T^{Msg1}	Processing time to detect a *Msg 1*	$2ms$ [10]
W^{Msg2}	*Msg 2* window	$5ms$ [10]
M^{Msg2}	Number of *Msg 1* in a single *Msg 2* message	6 [10]
U^{Msg2}	Maximum number of UEs receiving *Msg 2*	-
W^B	Backoff window size	$20ms$ [10]
S_i^E	Number of bearer/ACNB establishments	-
S_i^S	Number of bearer/ACNB re-setups	-
S_i^U	Number of bearer/ACNB updates	-

that is initialized to 1 when the UE triggers the RA procedure because of a new generated packet.

At the generic RA slot i, each UE $p \in \mathcal{A}_i$ sends a randomly selected *Msg 1*. The probability of successfully receiving the *Msg 2* depends on three aspects. Firstly, the success probability for *Msg 1* reception at the RAN, i.e., α, which is affected by the power ramping procedure used to transmit the *Msg 1* [10]:

$$\alpha(p, n_p) = 1 - \frac{1}{e^{n_p}} \tag{1}$$

The second component is the probability β of selecting a non-colliding *Msg 1*:

$$\beta(R, \mathcal{A}_i) = e^{-\frac{|\mathcal{A}_i|}{R}} \tag{2}$$

The third probability γ considers the limitations of resources at the RAN when transmitting *Msg 2*. Indeed, a single *Msg 2* message can list up to M^{Msg2} different *Msg 1*. The maximum number of UEs that can be acknowledged in a RA slot is thus given by $U^{Msg2} = M^{Msg2} \cdot W^{Msg2}$, where W^{Msg2} is the *Msg 2* window. The parameter γ can be computed as:

$$\gamma(R, \mathcal{A}_i) = \min \left(1, \frac{U^{Msg2}}{\beta(R, \mathcal{A}_i) \cdot |\mathcal{A}_i|} \right) \tag{3}$$

By considering the processing time T^{Msg1} for the RAN node to detect a *Msg 1*, the *Msg 2* window W^{Msg2}, a UE p that sent a preamble in the i-th RA slot (i.e., $p \in \mathcal{A}_i$) is expected to receive the *Msg 2* on average in the slot:

$$i^* = i + \left\lceil \frac{T^{Msg1} + W^{Msg2}/2}{T^{RA}} \right\rceil \tag{4}$$

In case W^{Msg2} expires without the reception of a *Msg 2*, the UE declares a failure in the RA attempt and schedules another

Msg 1 transmission ($n_p = n_p + 1$) with a backoff interval of $W^B + 1$. On average, the UE will thus perform a new attempt in the RA slot $i + \lceil \frac{(W^B+1)/2}{T^{RA}} \rceil$.

Let's indicate with $\mathcal{B}_{i^*} \subseteq \mathcal{A}_i$ the UEs receiving the *Msg 2* in the i^*-th RA slot. For a UE $p \in \mathcal{A}_i$, the probability of belonging to $\mathcal{B}_{i^*}$ is thus $\alpha(p, n_p) \cdot \beta(R, \mathcal{A}_i) \cdot \gamma(R, \mathcal{A}_i)$. Once a UE receives the *Msg 2*, it also receives a grant to transmit *Msg 3*. For simplicity, it is assumed that that the RAN node has enough resources for U^{Msg2} *Msg 3* during one RA slot. The set $\mathcal{C}_{i^*+1} = \mathcal{B}_{i^*}$ indicates the UEs transmitting a *Msg 3* in the RA slot following the one they received the *Msg 2*.

The reception of a *Msg 3* at the RAN node in a generic slot i triggers CP signaling for bearer/ACNB in the CN. For each UE $p \in \mathcal{C}_i$, the CN checks if p is already associated to a VD. Let's assume that there are no VDs currently activated that are associated to the same class c of UE p. The CN will then enable a new VD, v, (in a more general case, this means that the overall number of VDs will be increased, i.e., $V = V + 1$) and it will set $c_v = c | p \in \mathcal{P}_c$. As the VD has been initialized, i.e., $\mathcal{K}_v = \{p\}$, the CN will trigger a bearer/ACNB establishment. This is measured by updating S_i^E, which counts the number of establishments in the i-th RA slot.

If there are active VDs associated to the class c of the UE but p is not associated to any of those, the CN checks if there is a VD, v, that did not reach yet its maximum capacity, i.e., $v | p \in \mathcal{P}_{c_v}$, with $|\mathcal{K}_v| < U_{c_v}$. In case such VD, v, exists, the CN will add the UE p to the VD, i.e., $\mathcal{K}_v = \mathcal{K}_v \cup \{p\}$ and then the ACNB relevant to the v-th VD will be updated (updates are defined only for the solution exploiting the ACNB) to reflect the addition of the new UE. This is measured through the counter S_i^U, counting the number of ACNB updates[2] in the i-th RA slot.

When the CN is aware of the virtual device v the UE p is associated to, i.e., there exists a $v | p \in \mathcal{K}_v$, the CN will then check the lifetime T_{c_v} of the bearer/ACNB related to the v-th VD. If the timer is not expired, the CN will inform the RAN node to use the bearer/ACNB (already setup) associated to the v-th VD. In this case no CP traffic is needed. If T_{c_v} is expired, the CN will trigger a bearer/ACNB re-setup. This is measured through the counter S_i^S, counting the number of bearer/ACNB re-setup triggered in the i-th time slot.

In general, the CN will re-initialize the timer of the bearer/ACNB related to the the v-th VD any time there is UP traffic from a UE, p, belonging to the v-th VD.

V. PERFORMANCE EVALUATION

A. KPIs and Simulation Scenarios

This Section presents a quantitative evaluation of the benefits introduced by the connectivity solution for mIoT in enhanced 5G systems (hereinafter mIoT-5G-CM) compared to a baseline 4G connectivity model (hereinafter 4G-CM). In the conducted analysis, the UE density as well as their activation

period (AP) are varied. The evaluated key performance indicators (KPIs) are:

- *RAN-CN signaling*, i.e., measured as $S_i^E + S_i^U + S_i^S$. In 4G systems, it is the overall CP signaling from the eNB to the MME, in 5G systems from the AN to the AMF/SMF.
- *establishment* of bearers/ACNB, i.e., S_i^E.
- *re-setup* of bearers/ACNB, i.e., S_i^S.
- *update*, of ACNB, i.e., S_i^U.
- *UP resource utilization*, measured as $\sum_{i=1}^{I} |\mathcal{C}_i|/V$, i.e., the ratio between the UP traffic (under the assumption that for each *Msg 3* there is one UP packet) and the number of VDs activated to handle the P UEs.

The RAN is composed of ANs (ISD=1.732km) connected to a single CN. The RA parametrization, reported in Table I, is the same in 4G and 5G RANs.
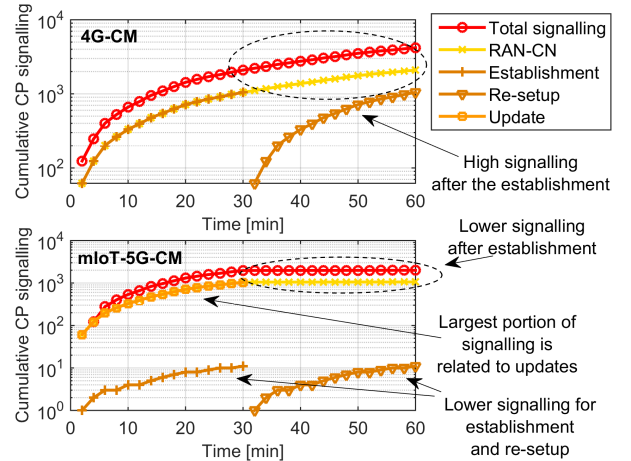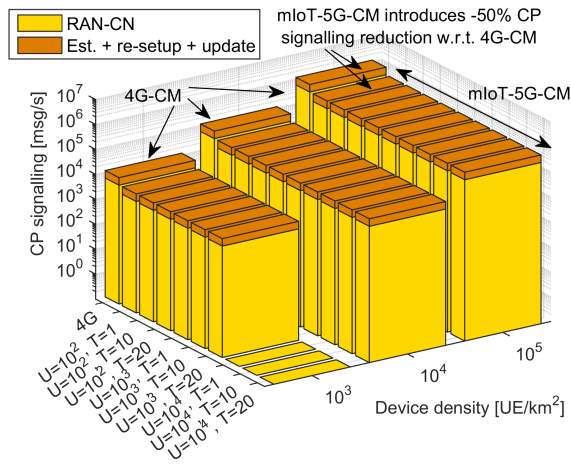

Fig. 3. Time variation of cumulative CP signaling during the interval $D = 1h$ (U=100, T=1, 10^3 UE/km^2, AP=30min).
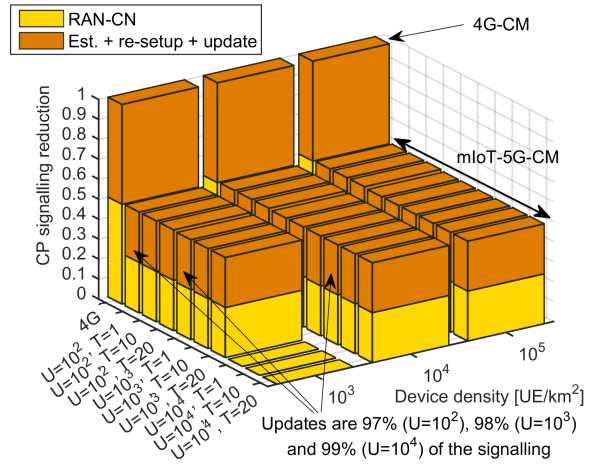
B. Simulation Results and Discussion

The performance analysis starts with an evaluation of the CP signaling during the interval time of interest, shown in Fig. 3. This analysis[3] aims at studying how the CP signaling from one AN evolves during the time as well as to understand how it is distributed among establishment, re-setup and updates (for mIoT-5G-CM). In 4G-CM, bearers are established during the first half of the interval D and it can be noticed that there is one establishment for each UE. With mIoT-5G-CM, establishments are reduced as performed on a ACNB-basis and the introduced reduction depends on the ratio between P and U. The higher portion of CP signaling is related to bearer updates (which involves less load in the CP compared to establishments). After the establishment, the CP signaling is related to re-setup. For 4G-CM, the load of re-setup is the same as in the establishment phase (i.e., one re-setup for each UE) and this means that the CP signaling towards the CN is always constant with 4G-CM. In mIoT-5G-CM, only one re-setup for each ACNB is needed. This means that after the establishment phase, the CP signaling

---

[2]Bearer/ACNB establishment and updates are considered in a different way as they could involve a different load for CP entities. The same holds for bearer/ACNB re-setup.

[3]UEs belong to the same class. The same holds for Fig. 4 and Fig. 5.

Fig. 4. CP signaling load (a) and CP signaling normalized to 4G-CM (b).

is limited to a maximum V re-setups (i.e., one re-setup for each VD).

After considering only one AN, the focus is now on the performance in a practical CN deployment with 10^4 ANs. Fig. 4(a) and Fig. 4(b) aim at studying the impact that device density, number of UE handled by a VD (U) and ACNB activity timer (T) have on mIoT-5G-CM. From this analysis, it emerges that the performance of this solution is not strongly affected by above listed parameters. In particular, in Fig. 4(b), the CP signaling is normalized to 4G-CM to better understand the introduced benefits w.r.t. 4G-CM as well as the impact of device density, U, and T. The reduction is 50% for this traffic pattern as AP=30min. UEs transmit on average two packets in the simulation period of 1 hour. It is interesting to note that the reduction of signaling does not significantly vary when increasing U or T. This is due to the fact that most of CP signaling for mIoT-5G-CM is related to updates, and this depends on the number of UEs.

When considering the UP resource utilization, shown in Fig. 5, it can be noticed the impact of the number of UEs handled by a VD (U) on this KPI. The higher U, the higher the number of UEs re-using the same ACNB and thus the higher the UP resource utilization. Although it may seem immediate that using higher U increases the UP resource utilization, it is worth to underline that higher U means higher number of UEs handled with the same ACNB thus higher UP resource request (i.e., bandwidth) for the related ACNB. This means that tuning U may depend on the available resources for the UP (i.e., network context): using higher U if enough UP resources are available or lower U on the contrary.

The above analyses considered that all UEs have the same traffic pattern. To consider a more realistic scenario, in Fig. 6
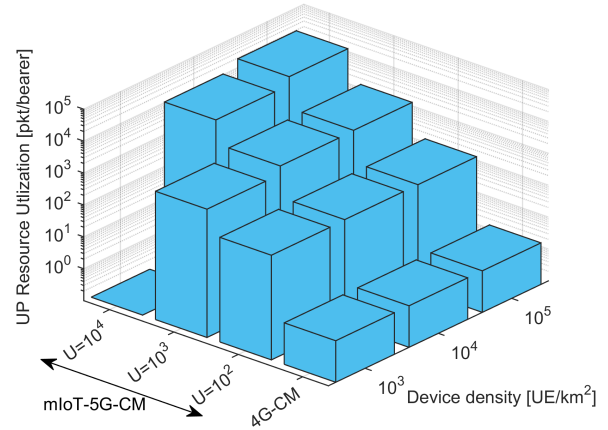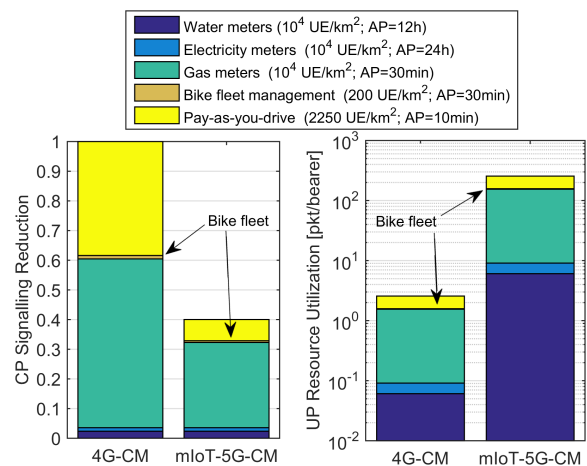

Fig. 5. UP resource utilization.

Fig. 6. CP signaling reduction and UP resource utilization in case of multiple classes (for all classes, U=100 and T=1min).
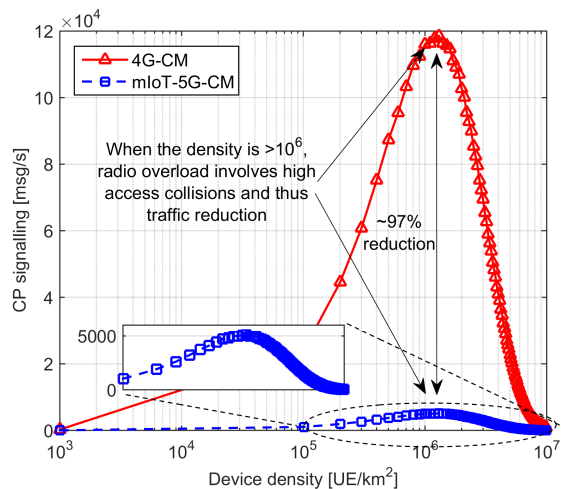
Fig. 7. CP signaling (same class for all UEs, AP=2min, U=100, T=1min) by varying the UE density to load the RAN node.

multiple types of traffic are considered[4]. The analysis shows that mIoT-5G-CM is able to guarantee a CP signaling reduction of 60% w.r.t. 4G-CM when considering a practical scenario with different types of active UEs with an increase in the UP resource utilization of 2 orders of magnitude.

Finally, in Fig. 7 the focus is on the behavior of the connectivity models for increased RAN load (to better isolate the behavior of the RAN, only one AN is considered). This analysis highlights that the CP signaling reaches a maximum value before starting the RAN congestion. When the UE density increases, the CP signaling goes close due to the huge amount of collisions in the RA procedure. Thus, a large portion of UEs is not able to send a connection request towards the CN. It is worth noticing that, when reaching the maximum load supported by the RAN (roughly 10^6 UE/km2 in the considered scenario), mIoT-5G-CM reduces the signaling towards the CN of about 97% (this is due to the fact that most of CP traffic is re-setup which is a heavy load for 4G-CM while mIoT-5G-CM drastically reduces this signaling thanks to the use of VD). The solution presented in this paper is thus able to avoid CP overload also in scenarios where the RAN is overloaded. This means that mIoT-5G-CM could allow RAN nodes to work close to their maximum congestion levels without affecting the CP entities in the CN.

VI. CONCLUSION

In this paper, the design and the performance evaluation of an E2E mIoT connectivity solution for 5G system has been presented. Starting from analysis of the envisaged device deployment scenarios and the expected traffic models, the presented solution targeted the requirements relating to massive device deployments and small data transmission. Assuming a 4G-like network deployment, the presented analysis aimed at

[4]Device density and AP of each traffic type are set as in [14] and are summarized in the legend of the plot. It is assumed that a ACNB gathers UEs belonging to the same traffic type.

highlighting the bottleneck at the CN for providing massive connectivity. Hence, the proposed E2E solution combined an mIoT enhanced LTE access with an mIoT 5GC featuring a new connectivity model. The new connectivity model is based on the Virtual Device and Aggregate Bearer concepts, together allowing the 5GC to handle multiple physical devices accessing the network via the same access node as a single logical element. The solution evaluation, still preliminary but considering relevant deployment scenarios and traffic models, has shown a 5GC CP signaling reduction up to 60% compared to a baseline LTE/EPC system when considering close-to-reality mIoT traffic, and the capability of dramatically improving the efficiency in UP resource utilization. Future work will focus on extending the performance evaluation to different mIoT device classes and diverse device deployment scenarios. Further device class specific optimizations will also be investigated.

REFERENCES

[1] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 510–527, March 2016.

[2] K. Mahmood, T. Mahmoodi, R. Trivisonno, A. Gavras, D. Trossen, and M. Liebsch, "On the integration of verticals through 5G control plane," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[3] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network slicing in 5G: An auction-based model," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[4] N. Zhang, P. Yang, S. Zhang, D. Chen, W. Zhuang, B. Liang, and X. S. Shen, "Software Defined Networking Enabled Wireless Network Virtualization: Challenges and Solutions," *IEEE Network*, vol. 31, no. 5, pp. 42–49, 2017.

[5] 3GPP, "System Architecture for the 5G System; (Release 15)." TS 23.501 V1.4.0, Sep 2017.

[6] 3GPP, "Procedures for the 5G System; Stage 2, (Release 15)." TS 23.502, v1.2.0, Sep 2017.

[7] 3GPP, "NR and NG-RAN Overall Description; Stage 2 (Release 15)." TS 38.300, v1.2.0, Oct 2017.

[8] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives," *IEEE Communications Surveys Tutorials*, vol. 16, pp. 4–16, First 2014.

[9] H. Thomsen, N. K. Pratas, C. Stefanović, and P. Popovski, "Code-expanded radio access protocol for machine-to-machine communications," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 355–365, 2013.

[10] 3GPP, "Technical Specification Group Radio Access Network;Study on RAN Improvements for Machine-type Communications; (Release 11)." TR 37.868 V11.0.0, Sep 2011.

[11] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13)." TR 45.820 V13.1.0, Nov 2011.

[12] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 15)." TS 23.401 V15.1.0, Sep 2017.

[13] 3GPP, "Feasibility Study on New Services and Markets Technology Enablers for Massive Internet of Things; Stage 1, (Release 14)." TR 22.861 V14.1.0, Sep 2016.

[14] Orange and Ericsson, "Traffic Model for legacy GPRS MTC; document GP 160060." 3GPP GERAN meeting #69, Feb 2016.