Comparison of Collision-Free and Contention-Based Radio Access Protocols for the Internet of Things

Marco Centenaro, *Student Member, IEEE*, Lorenzo Vangelista, *Senior Member, IEEE*, Stephan Saur, *Member, IEEE*, Andreas Weber, and Volker Braun

***Abstract*—The fifth-generation (5G) cellular networks will face the challenge of integrating the traditional broadband services with the Internet of Things (IoT), which is characterized by sporadic uplink transmissions of small data packets. Indeed, the access procedure of the previous generation cellular network (4G) is not able to support IoT traffic efficiently because it requires a large amount of signaling for the connection setup before the actual data transmission. In this context, we introduce two innovative radio access protocols for sporadic transmissions of small data packets, which are suitable for 5G networks because they provide a resource-efficient packet delivery exploiting a connection-less approach. The core of this paper resides in the derivation of an analytical framework to evaluate the performance of all the aforementioned protocols. The final goal is the comparison between 4G and 5G radio access solutions employing both our analytical framework and computer simulations. The performance evaluation results show the benefits of the protocols envisioned for 5G in terms of signaling overhead and access latency.**

***Index Terms*—Cellular Networks, Massive Radio Access, LTE, M2M, 5G, IoT, Mission Critical Communications, Random Access Protocols.**

I. Introduction

THE Internet of Things (IoT) is expected to play a fundamental role in improving the quality of our lives in near future, allowing the activation of new services that span from goods tracking to e-Health [3]. According to the Cisco® Visual Networking Index (VNI) Forecast [4], a huge growth of the Machine-to-Machine (M2M) market, which is the most important enabler of the IoT paradigm, is expected in the next five years. This implies that the number of Machine-Type Devices (MTDs), i.e., smart meters, wireless sensors, and actuators, will increase with an exponential trend. Nowadays, the majority of wireless technologies for M2M traffic are ad-hoc, short range wireless solutions, e.g., based on the IEEE 802.15.4 standard. However, in the perspective of a *place-&-play* notion, i.e., in a scenario in which an MTD needs just to be *placed* in the desired location to get connected to the rest of the world [5], the key role of cellular networks forcefully emerges. Indeed, terrestrial radio technologies are capable of providing ubiquitous coverage, with extremely low energy consumption, low complexity at the end device, possibly low latency, and minimal cost-per-bit. Nevertheless, current cellular standards are not originally designed to support this particular kind of traffic, therefore the IoT scenario needs to be considered as a major challenge for the next generation of wireless cellular systems, commonly referred to as fifth-generation (5G) [6]. The aim of this paper is proposing resource-efficient radio access protocols for the 5G air interface that are designed to support the demands of the IoT services. To evaluate the performance of the proposed solutions we derive a mathematical model and run empirical simulations, showing that a higher throughput and considerably lower delay with respect to the Long Term Evolution (LTE) standard can be achieved.

The rest of the paper is organized as follows. In Section II the air interface requirements to support IoT are presented, and the issues of current cellular network architectures are addressed. Then, in Section III the proposed radio access protocols are presented and compared with current cellular network standards. In Section IV we derive the analytical framework of the various radio access solutions and in Section V we provide the performance evaluation results. Finally, the conclusions are drawn in Section VI.

Manuscript received November 11, 2016; revised March 20, 2017; accepted May 13, 2017. This work is partly based on the conference papers [1], [2]. The associate editor coordinating the review of this paper and approving it for publication was V. Wong.

M. Centenaro and L. Vangelista are with the Department of Information Engineering, University of Padova, Italy; e-mail: marco.centenaro@dei.unipd.it, lorenzo.vangelista@unipd.it.

S. Saur, A. Weber, and V. Braun are with the Nokia Bell Labs, Stuttgart, Germany; e-mail: {stephan.saur, andreas.weber, volker.braun}@nokia.com.

II. Air Interface Requirements to Support IoT

In the last few years, Internet-based services like video streaming have become prominent for mobile users, thanks to more powerful User Equipments (UEs). To satisfy the needs of this kind of traffic, the latest cellular technologies, i.e., the Universal Mobile Telecommunication System (UMTS) [7] and the Long Term Evolution (LTE) [8], have been designed to provide the highest capacity available to a fairly limited amount of terminals. The IoT scenario, instead, entails a variety of new services, e.g., smart metering, building automation, e-Health, etc., requiring the sporadic uplink transmission of short messages, e.g., an Ethernet frame of 576 bits, to a remote server, where data analysis is performed. Therefore, the air interface of the cellular network in 5G needs to face the challenge of providing the highest capacity available to traditional terminals and, at the same time, ensure the network access to a big number of simultaneous transmissions coming from MTDs. Moreover, due to the great aggregate number of connected devices, the exchange of control information, both in uplink (UL) and downlink (DL), needs to be minimized. Finally, as MTDs are expected to be low-complexity, low-cost devices, the effort for establishing the radio access must be simplified at the terminal side, thus, shifting the burden to the base station, called eNodeB (eNB).

A. The Limits of Current Cellular Network Standards

Current cellular standards like the LTE were designed to provide broadband access to a reasonably small number of UEs, ideally directly proportional to the cardinality of the people inside the cell. When a terminal needs to connect to the network, it performs a Random Access (RA) procedure consisting of the following four steps.

1) *Preamble Transmission*. Each UE sends one signature (called *preamble*), randomly chosen among 64 ones,[1] to the eNB on the Physical Random Access Channel (PRACH). The preambles are mutually orthogonal, therefore the eNB is able to distinguish those UEs which choose distinct signatures. However, if more than one UE select the same preamble, we have a collision event that the eNB is usually not able to detect at this stage.[2]
2) *Random Access Response (RAR)*. For each detected request, the eNB transmits a response containing the time alignment command and Physical Uplink Shared Channel (PUSCH) resources assigned to the terminal. UEs that do not receive a RAR within a specific time interval must set a random backoff timer to start a new preamble transmission.
3) *Connection Request (CR)*. After the reception of the RAR, the UE sends a Radio Resource Control (RRC) message containing its unique Cell Radio Network Temporary Identifier (C-RNTI) on the indicated UL resources. Undetected collision events in the preamble transmission phase are revealed at this stage.
4) *Contention Resolution*. The eNB replies to those UEs whose CR was successful with a contention resolution message. Latest at this stage a collision of preambles is detected. The UEs that are not acknowledged by the eNB must restart the entire procedure after a random backoff interval.

After successfully completing the four-step RA procedure, further RRC signaling is needed before the UE is finally *connected* to the network. At this stage, a contention-free Scheduling Request (SR) opportunity on the Physical Uplink Control Channel (PUCCH) is assigned to each terminal to ask PUSCH resources for data transmissions.

The aforementioned LTE radio access procedure results to be inefficient in a M2M scenario for three distinct reasons:

a) as stated in [5], the massive number of preamble transmissions would cause the overload of RA procedure both in UL and DL due to the limited number of available signatures, thus, increasing collision probability, access delay, and failure rate;
b) moreover, additional DL resources would be also needed in presence of massive access requests, as the RAR message consists of 56 bits per UE;
c) finally, even assuming that a MTD succeeds in completing the access procedure, the uplink payload size is so small that the overall system efficiency would be significantly degraded due to the signaling overhead.

Therefore, we can state that the current LTE radio access procedure does not scale in the presence of massive access attempts by MTDs, resulting in a sharp degradation of the Quality of Service (QoS) of both conventional services and IoT services. With these considerations in mind, the authors propose an innovative, efficient, and flexible radio access protocol for IoT that complements the current LTE access procedure.

B. Radio Access Protocols for 5G: Related Work

Many solutions to provide network access to a massive number of terminals in wireless cellular networks have been proposed and discussed in literature [5], [10]. The 3rd Generation Partnership Project (3GPP), which is the standardization body responsible for most of the current cellular standards, presented some approaches to counteract the service degradation in case of overload of the PRACH due to massive MTD access requests, based mostly on the separation of the available resources [11]. Amendments of the LTE access procedure itself were proposed, as well, by 3GPP to decrease the RRC signaling in presence of M2M traffic [12]. Moreover, the Narrowband-IoT (NB-IoT) technology has been developed as a standalone standard to provide wide-area coverage to IoT terminals [13], but it is rather an ad-hoc solution, which is not capable of adapting to different scenarios. On the other hand, some studies in the literature suggested to improve the energy efficiency and the QoS of Machine-Type Communication by clustering the MTDs [14], [15] or employing game theoretic and machine learning approaches [16], [17]. In [18] a Self-Optimizing Overload Control (SOOC) is proposed that can be combined with existing LTE overload mechanisms. It can tune random access resources according to the load situation. However, a reduction of the signaling overhead is not possible. Finally, several *clean slate* approaches were designed to try to solve the problem at its roots. For example, in [19] authors propose that MTDs with different priorities first contend for network access with a p-persistent Carrier Sense Multiple Access (CSMA) mechanism; then, successful devices are assigned a time slot for transmission in a Time Division Multiple Access (TDMA) fashion. In [20] the paradigm of *coded random access*, in which the structure of the access protocol can be mapped to a structure of an erasure-correcting code defined on a graph is described. In [21] Compressive Sensing-based Multi-User Detection (CS-MUD) techniques are combined with multicarrier access schemes to increase the spectral efficiency and reduce the control signaling overhead and processing complexity required to handle a massive number of MTDs. In [22] authors derive an entire frame design for a low-complexity Time Division Duplex (TDD) system with suitable radio numerology for massive connection density and bursty packet transmissions.

The fundamental drawback of the majority of these solutions is that they require a brand-new, additional air interface, which should be *separated* from the current cellular network interface, thus their implementation would be a serious issue.

[1]The actual number of available signatures is typically lower, usually equal to 54, because some of them are reserved for special purposes.

[2]The eNB may not detect a collision in step 1 due to the capture effect of the channel or because the collided UE are not separable in terms of Power Delay Profile (PDP) [9]. For this reason, in practical systems the detection of collided preambles is often not considered, thus in the following of this paper we assume that the eNB is not able to detect a collision event at step 1.

Moreover, the additional air interface could not be used for other services in case IoT traffic is not present, thus wasting the allocated spectrum. The aim of our work, instead, is to provide a *unified* air interface for 5G which is able to integrate broadband services and IoT services at the same time, ensuring the backward compatibility with current cellular standards like LTE. With this flexible design, radio resources can be dynamically allocated for those services that actually need them. Simultaneous support of multiple services by sub-band wise optimization of physical layer (PHY) parameters like subcarrier spacing or pulse shape is discussed in [23] and [24], respectively. An additional filtering of the sub-bands can mitigate Inter-Service-Band-Interference (ISBI) [25].

III. Proposed Radio Access Protocol

In this section the resource-efficient radio access schemes for IoT terminals are presented. Firstly, the PHY specifications are described and, then, the proposed solution is introduced in two variants, i.e., the *One-Stage* protocol and the *Two-Stage* protocol. Possible feedback formats are discussed and, finally, a comparison with LTE is provided. Without loss of generality, in the following we assume perfect synchronization of all UL transmissions at the eNB.

A. Physical Layer Design

Let us refer to Fig. 1 and consider a multi-carrier transmission system, based on an Orthogonal Frequency Division Multiple Access (OFDMA), consisting of elementary resource units called Resource Elements (REs), equivalent to one subcarrier and one time symbol (OFDM symbol). A group of REs over S subcarriers and T symbols forms a Resource Block (RB). In the following we assume that a RB spans a period of one subframe, also called Transmission Time Interval (TTI), of duration T_{TTI}. We remark that such a PHY design is implemented by the latest cellular network technologies like LTE.

Without loss of generality, we define a Small Packet Block (SPB) as a group of J RBs stacked in frequency. The time duration of a SPB is still one subframe, as for the RB. In every subframe, M SPBs are available, where M_{SR} SPBs are dedicated for scheduling requests and M_{D} SPBs for actual data transmission,[3] in such a way that

$$M = M_{\mathrm{SR}} + M_{\mathrm{D}} \,. \tag{1}$$

In the proposed solution a SR is represented by a code sequence, i.e., a signature, similar to an LTE preamble, that is mapped on the radio resources. If we assume that R orthogonal codes can be distinguished per RB, a total amount of RJM_{SR} SRs per TTI can be detected at the eNB side. There is not a one-to-one mapping between SRs and data SPBs, but we rather allow for an *over-provisioning* of SRs, i.e., we assume that the number of signatures is greater than the actually available data resources. For this reason, the over-provisioning parameter N is introduced, so that

$$R \times J \times M_{\mathrm{SR}} \simeq N \times M_{\mathrm{D}} \,. \tag{2}$$

Note that parameter N may equivalently be defined as the ratio between the aggregate number of SRs and the amount of data SPBs. Substituting $M_{\mathrm{D}} = M - M_{\mathrm{SR}}$ from (1) in (2), the value of M_{SR} can be determined as a function of N, yielding

$$M_{\mathrm{SR}} = \left\lceil \frac{NM}{RJ+N} \right\rceil . \tag{3}$$

Equivalently, N can be obtained as a function of M_{SR} as follows:

$$N = \left\lfloor \frac{RJM_{\mathrm{SR}}}{M - M_{\mathrm{SR}}} \right\rfloor . \tag{4}$$

Note that $N \geq 1$, since at least one signatures should be associated to every data SPB. We remark that, without loss of generality, we assume N to be an integer value in this derivation and in the following of the paper. A graphical representation of the SR mapping into RBs and of the over-provisioning factor N with respect to one data SPB is provided in Fig. 2.

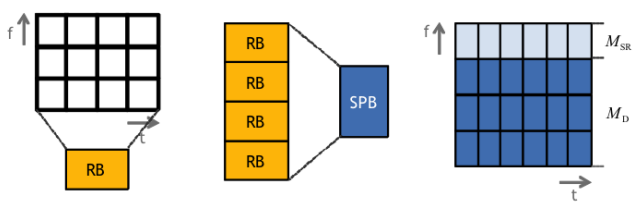

Fig. 1. Proposed OFDM structure. The white boxes denote the REs, the yellow ones the RBs, and the blue ones the SPBs.

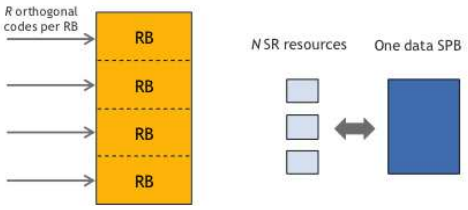

Fig. 2. Mapping of SR resources in RBs and over-provisioning factor N compared to data SPBs.

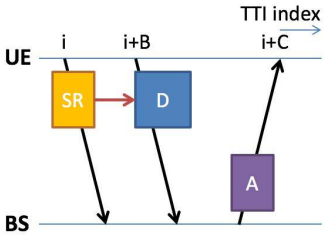

Fig. 3. One-Stage Protocol.

[3]As an alternative to this Frequency Division Multiplexing (FDM) scheme, which complies to the current resource allocation scheme in LTE, a Time Division Multiplexing (TDM) scheme can be applied in the case of a very small system bandwidth. However, as access delay is one key performance indicator, in the following of the paper we will mainly focus on FDM. A hybrid FDM/TDM scheme will be proposed in Sec. III-D2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2707074, IEEE Transactions on Communications

B. One-Stage Protocol

The first variant of the proposed solution is called *One-Stage* protocol. A graphical representation of the protocol can be found in Fig. 3. The protocol consists of three steps.

1) The MTD sends a SR with index $r = 1, \ldots, NM_{\mathrm{D}}$ that points uniquely to one specific payload resource.
2) The MTD sends its data packet utilizing the SPB corresponding to the chosen SR, either in the same subframe or in one of the subsequent subframes.
3) If the data transmission is successful, then the eNB acknowledges the packet; otherwise a not-acknowledgement (NACK) message is sent to the MTD.

The transmission of the SR in step 1 is utilized for *activity detection*. Although the reservation of extra RBs for SRs is not mandatory if the One-Stage protocol operates in a standalone system, we remark that step 1 becomes necessary in a multi-service interface that has to support both collision-free and contention-based data transmissions. Furthermore, the SR can implicitly hold some extra information like the size of the SPB or the used Modulation and Coding Scheme (MCS) for the data. Finally, we observe that the predefined mapping between SRs and data SPBs could be disadvantageous in the presence of frequency selective fading. In case the MTD has some channel knowledge based previous transmission attempts, which is a valid assumption if the propagation conditions are static, the MTD can avoid SRs pointing to data resources in a deep fade.

This radio access solution is much faster than LTE in case the transmission is successful, and it requires significantly less DL feedback. There are some disadvantages, though: the high collision probability reduces the throughput and the coexistence with scheduled transmissions may be difficult to handle. For these reasons, this solution is envisaged for very small packets and low traffic load.

C. Two-Stage Protocol

The benefit of a high over-provisioning factor N resides primarily in a reduction of the probability that more than one MTDs select the same preamble index to send a SR. Nevertheless, this positive effect is not exploited in the One-Stage protocol. Indeed, even though the terminals pick different preambles, if their SRs point to the same data SPB, we cannot avoid the collision in step 2 and all collided data packets are lost. Furthermore, to provide a higher value of N we have to increase M_{SR} and, consequently, decrease M_{D}. Therefore, the best option for the One-Stage approach is to minimize the value of N.

On the other hand, we may take advantage of over-provisioning as follows. As one variant of the previous protocol let us assume that the eNB is able to reject part of the detected SRs in order to prevent packet collisions on data SPBs.[4] This second proposed approach is called *Two-Stage* protocol and its operation is graphically explained in Fig. 4. The protocol consists of four steps.

[4]Alternatively, over-provisioning of SRs could be exploited through Multi-User Detection (MUD) techniques. If the eNB is aware that one data SPB is utilized by multiple UEs, it may apply, e.g., Successive Interference Cancelation (SIC). This aspects will be part of our future work.

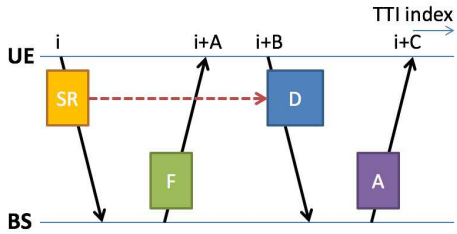

Fig. 4. Two-Stage protocol.

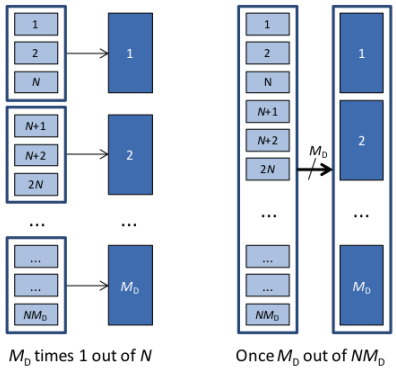

Fig. 5. Difference between tagged data SPBs (left hand side) and pooled data SPBs (right hand side).

1) The MTD chooses a random index $r = 1, \ldots, NM_{\mathrm{D}}$ and sends a SR to the eNB, requesting UL radio resources.
2) The eNB sends information related to the assignment of a data SPB as feedback, i.e., which SPB to use and in which subframe to use it. In case the SR is not acknowledged, the MTD randomly selects a new index r and sends a new SR after a random time offset $\beta \in [\beta_{\min}\ ,\ \beta_{\max}]$.
3) The MTD sends data utilizing the assigned data SPB.
4) The eNB acknowledges the data transmission if received successfully. In case the data SPB is not acknowledged, the MTD restarts the procedure from step 1. The available number of SR transmissions are restricted to Θ in order to avoid the overload of the system.

This second radio access solution is promising because, using a high over-provisioning factor, it reduces the collision probability and, therefore, increases the throughput. Of course, collisions on data resources still happen if more than one MTD choose the same preamble index in step 1. On the other hand, it requires an additional delay with respect to the One-Stage protocol due to the feedback required after the SR transmission. For these reasons, this solution is envisaged for bigger packets and high traffic load.

The authors remark that a dynamic usage of the two protocols is possible, according to the traffic load.

Physical Resources Grouping: We can further customize the Two-Stage approach by splitting the total amount of available data SPBs M_{D} into K distinct groups, where K is such that $1 \leq K \leq M_{\mathrm{D}}$. This splitting can be either fixed per specification or can be adapted dynamically by the eNB per DL control channel according to the current traffic situation. For the sake of simplicity, in the following we assume that every group

comprises the same number of SPBs M_{D}/K and has the same arrival rate of new users. Consequently, $NM_{\mathrm{D}}/K \simeq RJM_{\mathrm{D}}/K$ SRs are associated to every group, and each MTD sends a SR that is associated with the targeted group. Two special cases are noteworthy:

1) the *tagged data* case, in which $K = M_{\mathrm{D}}$, i.e., each group consists of exactly one SPB,
2) and the *pooled data* case, in which, instead, $K = 1$, i.e., all SPBs belong to one single group.

A graphical representation of these extreme cases can be found in Fig. 5. In the first special case, each SR points exactly to one single data resource. As a consequence, the required feedback from the eNB is minimized (just acknowledgement (ACK) or NACK must be indicated, since the data SPB is already fixed), however, the scheduling is not flexible. With the transmission of the SR, it is already clear which SPB will be utilized for the data packet later on. As for frequency selective channels, similar consideration to the One-Stage case can be done. In the pooled case, instead, there is no predefined tagging between SRs and data SPBs. Consequently, the eNB has full scheduling flexibility, i.e., the eNB can distribute the full set of SPBs among the received SRs. This comes along with the drawback of an increased feedback effort in the DL: for each acknowledged SR the eNB has to indicate the assigned data SPB or vice versa.

The physical resource grouping allows for a differentiation of service types, device classes, packet sizes, or transport formats in a real and complex communication system. As an example, $K = 2$ groups can be configured, one group for high priority services (e.g. fire alarms), which consists of a big number of SPBs for a comparably low number of MTDs, and, vice versa, a second group for low priority services (e.g. air temperature measurement), which consists of only few SPBs for many MTDs. Consequently, the high priority service will experience a significantly lower collision probability and a higher success rate.

D. Feedback Formats

As DL signaling efficiency concerns, we provide some consideration about the feedback format for the proposed protocols, focusing both on the case of a fixed time relation and a relaxed time relation between the steps of the protocol.

1) Fixed Time Relation: Let us assume that a fixed time relation exists between the steps of the Two-Stage protocol, e.g., A, B, C TTIs as depicted in Fig. 4. As a consequence, the SR feedback from the eNB consists only of the particular SPB index that is assigned to each request. The following feedback formats for the SR ACK are proposed.

- *Option 1*: For every SPB the acknowledged SR is indicated. Since for every SPB we have to identify the SR we acknowledge within the corresponding group, we need a binary vector of length

$$F_1 = M_{\mathrm{D}}\left\lceil \log_2\left(\frac{NM_{\mathrm{D}}}{K}\right)\right\rceil [\mathrm{bit}]\,. \tag{5}$$

  Note that the number of required bits for this feedback format is very low, but the MTD must make M_{D}/K searches of the SR it sent.

- *Option 2*: For every SR the assigned data SPB is indicated. Since for every SR we have to identify the assigned SPB within the corresponding group, we need a binary vector of length

$$F_2 = NM_{\mathrm{D}}\left\lceil \log_2\left(\frac{M_{\mathrm{D}}}{K}+1\right)\right\rceil [\mathrm{bit}]\,. \tag{6}$$

  Note that the +1 accounts for an additional codeword for the NACK. Moreover, in the case of tagged data resources ($K = M_{\mathrm{D}}$) it is $F_2 = NM_{\mathrm{D}}$. We remark that this kind of option is larger in terms of bits, but the UE now does not need to search for the SR it sent.

2) Relaxed Time Relation: A performance gain is expected if the constraint of fixed time scheduling is relaxed, i.e., if we allow to delay a data packet transmission from an entirely occupied subframe to a later one. Two approaches for a fully flexible data SPB scheduling are proposed.

a) Feedback with Time Stamp: Considering a window of W subframes, we assume that the feedback comprises, in addition to the assigned SPB, a subframe index $w = 0, \ldots, W-1$ indicating the additional delay that has to be added to the minimal offset between the reception of SR feedback and the data transmission. Note that the fixed time relation is a particular case of the relaxed time relation with $W = 1$. Under this assumption, the length of feedback messages are

$$F_1^{(\mathrm{TS})} = M_{\mathrm{D}}\left\lceil \log_2\left(\frac{WNM_{\mathrm{D}}}{K}\right)\right\rceil [\mathrm{bit}] \tag{7}$$

using Option 1 and

$$F_2^{(\mathrm{TS})} = NM_{\mathrm{D}}\left\lceil \log_2\left(\frac{WM_{\mathrm{D}}}{K}+1\right)\right\rceil [\mathrm{bit}] \tag{8}$$

using Option 2. A graphical example of time stamp feedback is provided in Fig. 6, assuming that $M_{\mathrm{D}}/K = 2$, $N = 3$, $W = 2$, $A = 4$, and $B = 8$. In subframe i, 3 MTDs choose indices $r_1 = 2$, $r_2 = 4$, and $r_3 = 5$, respectively, and send their SRs. The third MTD, after the default delay of $A = 4$ TTIs, reads the feedback information, but does not find the acknowledgement of $r_3 = 5$. Since $W = 2$, the MTD is allowed to look for its SR again in subframe $i + A + 1 = i + 5$. As it finds its SR in the feedback together with $w = 1$, it starts the data transmission on SPB number 1 in subframe $i + B + w = i + 9$. Note that, if the third MTD had found $w = 0$, it would have not interpreted the feedback in subframe $i + 5$ as for itself, but as the acknowledgement for another UE that sent the same SR in subframe $i + 1$.

b) Queueing-Based Feedback: A promising approach to reduce the number of required feedback bits in the Two-Stage protocol with pooled resources consists in the Distributed Queueing Random Access Protocol (DQRAP) [26]. Many DQRAP-based protocols have been designed for wireless communications, e.g., for Wireless Local Area Network (WLAN) [27]. The drawback of this approaches is, however, that the eNB needs to be able to detect the collision of MTDs utilizing the same SR resource. This would require an immense additional complexity at the eNB receiver side. Therefore, a simplified queueing scheme using a single queue instead of two, as in DQRAP, has been designed. The queueing-based feedback

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2707074, IEEE Transactions on Communications

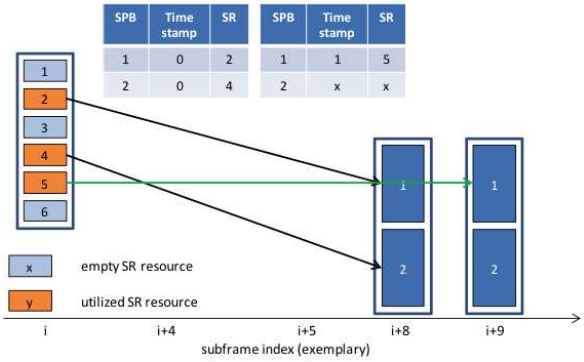

Fig. 6. Example of time stamp feedback with $M_{\mathrm{D}}/K = 2$, $N = 3$, $W = 2$, $A = 4$, and $B = 8$.

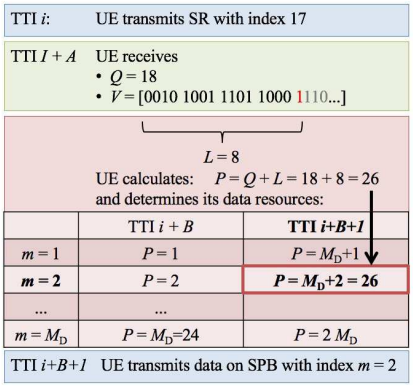

Fig. 7. Example of queueing-based feedback, assuming $M_{\mathrm{D}} = 24$ and $K = 1$.

TABLE I
MAPPING BETWEEN POINTER $P = Q + L$ AND TIME-FREQUENCY POSITION OF ASSIGNED SPB FOR DATA TRANSMISSION.

	$i+B$	$i+B+1$	$i+B+2$
$m=1$	$P=1$	$P=M_{\mathrm{D}}+1$	$P=2M_{\mathrm{D}}+1$
$m=2$	$P=2$	$P=M_{\mathrm{D}}+2$	$P=2M_{\mathrm{D}}+2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m=M_{\mathrm{D}}$	$P=M_{\mathrm{D}}$	$P=2M_{\mathrm{D}}$	$P=3M_{\mathrm{D}}$

consists of the length Q of queue $\mathcal{Q}$, accounting for the number of terminals that have already been acknowledged and are waiting to transmit, and of binary vector V, indicating for every SR whether it is active ($v_l = 1$) or not ($v_l = 0$), where v_l is the l-th value in vector V. Upon receiving this kind of feedback, a UE that chooses the SR index $r = X$ computes pointer P as follows:

$$P = Q + \sum_{l=1}^{X} v_l = Q + L, \tag{9}$$

where $L \triangleq \sum_{l=1}^{X} v_l$. The assigned TTI index t and SPB index m are derived from P according to Tab. I, where it is assumed that the minimal timing offset between SR and data transmission is of B subframes. It can be seen that $t = i + B + \lfloor (P-1)/M_{\mathrm{D}} \rfloor$, while $m = [(P-1) \mod M_{\mathrm{D}}] + 1$.

An example of queueing-based feedback is provided in Fig. 7, where we assume $M_{\mathrm{D}} = 24$, $K = 1$, and $B = 8$. We assume that a MTD chooses index $r = 17$ and sends the SR to the eNB. In subframe $i + A$ the MTD receives as feedback information $Q = 18$ and the vector V that is shown in the figure. It computes $L = 8$ and $P = 18 + 8 = 26$ and realizes that it has been scheduled in subframe $i + B + 1 = 9$ in SPB $m = 2$.

Note that the feedback message length using the queueing-based feedback is $F_Q = NM_{\mathrm{D}} + \lceil \log_2 Q \rceil$ [bit], where NM_{D} is the length of vector V. However, this feedback scheme can be generalized to $K > 1$ as well, sending K different values of Q and the vector V. Therefore, we can generalize the feedback length as follows:

$$F_Q = K\left(\frac{NM_{\mathrm{D}}}{K} + \lceil \log_2 Q \rceil\right) = NM_{\mathrm{D}} + K\lceil \log_2 Q \rceil \,[\mathrm{bit}]. \tag{10}$$

3) Feedback Comparison: A comparison of all the proposed feedback formats is provided in Tab. II, assuming $M_{\mathrm{D}} = 24$, $N = 3$, and $W = 9$. To provide a fair comparison between the approaches with a relaxed time relation, we must provide a conversion between the window size W and the queue length Q. Assuming the *worst* case for the queueing-based approach, in which the generic MTD points the last data SPB, it is:

$$W \geq \frac{\max\{P\}}{\frac{M_{\mathrm{D}}}{K}} = \frac{Q + \frac{M_{\mathrm{D}}}{K}}{\frac{M_{\mathrm{D}}}{K}} = \frac{K}{M_{\mathrm{D}}}Q + 1 \tag{11}$$

yielding

$$W = \left\lceil \frac{K}{M_{\mathrm{D}}}Q + 1 \right\rceil \quad \text{and} \quad Q = \left\lfloor (W-1)\frac{M_{\mathrm{D}}}{K} \right\rfloor. \tag{12}$$

We infer that in the case of tagged resources the most efficient feedback format is option 1 with time stamp, while in case of pooled resources the most efficient one is the queueing-based feedback. It must be taken into account, however, that option 1 with time stamp forces the MTD to look for its SR in the feedback message, i.e., it is more computationally demanding with respect to the other formats.

E. Comparison with LTE

The proposed radio access solution provides many advantages with respect to LTE. Firstly, it introduces a *contention-based* transmission paradigm in which a MTD is allowed to send UL data without undergoing the LTE four-step access procedure for collision resolution, thus reducing the signaling overhead, in particular the DL feedback, decreasing the packet delivery delay, and allowing for a significantly higher number of simultaneously active MTDs per radio cell. Collision resolution is achieved through retransmissions after a random backoff time, which is sufficient for most delay-tolerant applications. The proposed feedback formats, indeed, are *broadcast* messages, while the LTE RAR consists of individual messages, since time offsets for each MTD are included and resources for Message 3 are not pre-configured. In particular, according to [28], every RAR requires one octet for the header and six octets for the data, i.e., 56 bits per SPB. Moreover, the configuration of Physical Downlink Shared Channel (PDSCH), which carries the RAR messages, requires a Downlink Control

TABLE II
FEEDBACK FORMAT LENGTHS FOR THE TWO-STAGE PROTOCOL, ASSUMING $M_{\mathrm{D}} = 24$, $N = 3$, AND $W = 9$.

FEEDBACK FORMAT	GENERAL FORMULA [BIT]	2-STAGE TAGGED ($Q = 8$)	2-STAGE POOLED ($Q = 192$)
Option 1	$M_{\mathrm{D}}\left\lceil \log_2\left(\frac{NM_{\mathrm{D}}}{K}\right)\right\rceil$	**48**	**168**
Option 2	$NM_{\mathrm{D}}\left\lceil \log_2\left(\frac{M_{\mathrm{D}}}{K}+1\right)\right\rceil$	72	360
Option 1 with time stamp	$M_{\mathrm{D}}\left\lceil \log_2\left(\frac{WNM_{\mathrm{D}}}{K}\right)\right\rceil$	**120**	240
Option 2 with time stamp	$NM_{\mathrm{D}}\left\lceil \log_2\left(\frac{WM_{\mathrm{D}}}{K}+1\right)\right\rceil$	288	576
Queueing based	$NM_{\mathrm{D}} + K\lceil \log_2 Q\rceil$	144	**80**

TABLE III
COMPARISON OF FEEDBACK LENGTHS AFTER SR TRANSMISSION FOR LTE AND THE PROPOSED PROTOCOL. FOR THE TAGGED CASE AND THE POOLED CASE, THE OPTION 1 WITH TIME STAMP AND THE QUEUEING-BASED FEEDBACK FORMATS ARE CONSIDERED, RESPECTIVELY.

	LTE	1-STAGE	2-STAGE TAGGED ($Q = 8$)	2-STAGE POOLED ($Q = 192$)
GENERAL FORMULA [BIT]	$M_{\mathrm{D}}(8+48)$	0	$M_{\mathrm{D}}\lceil \log_2(WN)\rceil$	$NM_{\mathrm{D}} + \lceil \log_2 Q\rceil$
$M_{\mathrm{D}} = 24$, $N = 3$, $W = 9$	**1344**	0	120	80

Information (DCI) Format 1A message of 25 bits [29]. It is worth noticing that PHY overhead should be accounted also for the proposed feedback format, but 5G PHY specifications are not yet defined. However, assuming that 5G physical channels will be similar to LTE ones and considering that in LTE the Medium Access Control (MAC) overhead is predominant, we neglect the PHY overhead contribution and compare just the MAC layer overhead. As shown in Tab. III, a number of 24 SRs would result in an aggregate RAR size of 1344 bits including headers, thus the proposed feedback formats provide an improvement of more than one order of magnitude with respect to LTE.

Secondly, the proposed protocols can be easily combined with the *connectionless* concept [30]. As machine-type traffic is characterized by *sporadic infrequent* transmissions of small packets, the *connection-oriented* paradigm of LTE is highly inefficient. Apart from the four-step RA protocol itself, a cascade of signaling messages has to be exchanged between the MTD and the network before the MTD is in `RRC_CONNECTED`, `IN_SYNC` state and data transmission is possible. We aim to avoid this overhead and include source and destination addresses directly in the SPB without prior connection setup, as recommended in [31]. Apart from a more efficient usage of radio resources, the connectionless concept helps to reduce energy consumption in the MTD, mainly due to a much shorter on-time of the radio module in the MTD compared to LTE. This helps to achieve a clearly longer battery lifetime: sensor networks, indeed, typically aim for a MTD battery lifetime of several years, but LTE has not been designed for that purpose. We remark that other solutions were proposed in literature in this direction [32], [33], but their focus is rather on a mere re-engineering of LTE RRC procedures to support M2M traffic, whereas our contribution offers an additional degree of freedom with respect to RRC states. Also, we complement these concepts with a pervasive mathematical framework that can be generally applied for various configurations of radio access protocols.

Finally, the authors remark that the protocol implementation is not an issue, because current LTE physical channel specifications may be partly reused, but also selectively complemented by novel concepts like Universal Filtered OFDM (UF-OFDM), also known as Universal-Filtered Multi-Carrier (UFMC), that allows for relaxed synchronization for data transmission without severe performance degradation [34]. Filter Bank Multi-Carrier (FBMC) [35] and Filtered-OFDM (F-OFDM) [36] are alternative waveform concepts addressing the same problem: the common approach is the application of filtering to minimize the mutual interference between users caused by imperfect synchronization of UL signals.

IV. MATHEMATICAL MODEL

In this section we propose a mathematical characterization of the proposed radio access protocols using an analytical approach similar to [37]. We assume that the arrival process of new packets at the system follows a Poisson distribution of rate λ (expressed in packets per second). The overall arrival rate at the system, denoted with λ_T, is obtained summing new transmission attempts and retransmissions, i.e., $\lambda_T = \lambda + \lambda_R$. It is assumed also that the time interval between two Random Access Opportunities (RAOs), denoted as δ_{RAO}, is equal to one TTI, since in every subframe M_{SR} SPBs are available for scheduling requests.

A. Model of the One-Stage Protocol

From the perspective of a generic device in a set of j nodes, each of which randomly chooses one resource out of n available

resources, the probability that another contender node selects the same resource is

$$q(j,n) \triangleq 1 - \left(1 - \frac{1}{n}\right)^{j-1} . \tag{13}$$

Let us define, then, the one-shot[5] *failure probability* p_f as the average of $q(j,n)$, with $n = M_{\mathrm{D}}$, over the Poisson distribution of j overall arrivals at the system in one RAO:

$$\begin{aligned} p_f &= \mathbb{E}_j[q(j, M_{\mathrm{D}})] \\ &= \sum_{j=1}^{+\infty} \left[1 - \left(1 - \frac{1}{M_{\mathrm{D}}}\right)^{j-1}\right] \times e^{-\Delta} \frac{\Delta^j}{j!} \leq 1 - \left(1 - \frac{1}{M_{\mathrm{D}}}\right)^{\Delta-1} , \end{aligned} \tag{14}$$

where $\mathbb{E}[\cdot]$ denotes the expected value and $\Delta \triangleq \lambda_T \delta_{\mathrm{RAO}}$. We remark that the inequality holds for the Jensen's inequality and p_f is a function of λ_T.

The impact of multiple transmission attempts can be evaluated as presented in [37] by exploiting the Bianchi's model [38]. Recalling that Θ denotes the maximum number of transmission attempts, it can be proved that the *outage probability*, i.e., the probability of exceeding the maximum number of transmission attempts, of the radio access protocol is given by

$$p_{\mathrm{outage}} = p_f^{\Theta} . \tag{15}$$

The *average number of transmission attempts* is

$$\bar{\theta} = \sum_{\theta=1}^{\Theta} \theta \times \mathbb{P}[\theta \text{ tx}] = \sum_{\theta=1}^{\Theta-1} \theta p_f^{\theta-1}(1-p_f) + \Theta p_f^{\Theta-1} = \frac{1-p_f^{\Theta}}{1-p_f} , \tag{16}$$

where $\mathbb{P}[\theta \text{ tx}]$ is the probability that a packet undergoes θ transmission attempts. If we count only successfully delivered packets the mean number of transmission attempts becomes

$$\begin{aligned} \bar{\theta}_{\mathrm{ACK}} &= \sum_{\theta=1}^{\Theta} \theta \times \mathbb{P}[\theta \text{ tx}|\text{pkt ok}] = \\ &= \sum_{\theta=1}^{\Theta} \theta \times \frac{p_f^{\theta-1}(1-p_f)}{1-p_{\mathrm{outage}}} = \frac{1-(\Theta+1)p_f^{\Theta}+\Theta p_f^{\Theta+1}}{(1-p_f)(1-p_f^{\Theta})} , \end{aligned} \tag{17}$$

where $\mathbb{P}[\theta \text{ tx}|\text{pkt ok}]$ is the probability that a packet undergoes θ transmission attempts given that it is successfully delivered.

Finally, the value of λ_T can be determined solving the following fixed-point equation:

$$\lambda_T = \bar{\theta} \times \lambda = \frac{1-p_f^{\Theta}}{1-p_f} \times \lambda . \tag{18}$$

Note that if $\Theta = 1$ then $\lambda_T = \lambda$. The *throughput*, defined as the number of successful data packets per overall number of SPBs, can then be computed as

$$\mathcal{S} = \lambda \times (1 - p_{\mathrm{outage}}) . \tag{19}$$

[5]Allowing just one transmission attempt.

B. Model of the Two-Stage Protocol with Pooled Resources

Let us split the analysis of the protocol in two phases: the *preamble transmission* phase and the *data transmission* phase. The one-shot failure probability in this case is defined as follows:

$$p_f \triangleq 1 - (1-p_c)(1-p_d^A), \tag{20}$$

where p_c is the *collision probability* in the preamble transmission phase and p_d is the *dropping probability* during the access granted phase.

The collision probability can be computed similarly to Eq. (14), simply considering now the number of preambles in place of the amount of data SPBs. Therefore, we obtain

$$p_c = \mathbb{E}_j[q(j, NM_{\mathrm{D}})] \leq 1 - \left(1 - \frac{1}{NM_{\mathrm{D}}}\right)^{\Delta-1} . \tag{21}$$

The data transmission phase, instead, is modeled as a *queueing system* in which the customers, i.e., the successful SRs, are *impatient* customers [39]. In particular, we are interested in evaluating the *long-term fraction of users who are lost*, that is, the dropping probability of the queue. Let us denote the arrival rate at the queue, the queue service rate, the number of servers, and the maximum waiting time with Λ, μ, m, and τ, respectively. The dropping probability for a M/M/m queue is defined as

$$p_d(\Lambda, m\mu, \tau) \triangleq \frac{(1-\rho)\rho\Omega}{1-\rho^2\Omega} , \tag{22}$$

where $\rho = \Lambda/(m\mu)$ and $\Omega = e^{-m\mu(1-\rho)\tau}$. The system should be modeled as a M/D/m queue with impatient costumers, but no closed-form expression is known for this kind of queues. Nevertheless, according to [40], the expression of dropping probability for M/M/m queues is an excellent approximation for M/G/m queueing systems, including M/D/m queueing systems.

In the Two-Stage protocol, the arrival rate at the queue, denoted by λ_A, is the number of activated preambles (both collided and not) per time unit and it can be computed as follows. Let us define the random variable X as the number of users selecting the same preamble index. Since the average number of arrivals per preamble per subframe is $\alpha = \Delta/(NM_{\mathrm{D}})$, X is distributed according to a Poisson distribution of parameter α. We denote with ω the probability of preamble activation (a function of parameter α),

$$\omega(\alpha) = 1 - \mathbb{P}[X = 0] = 1 - e^{-\alpha} , \tag{23}$$

and assume that preamble activations are independent from each other. Then, we can compute the average arrival rate at the access granted λ_A as

$$\lambda_A = \frac{NM_{\mathrm{D}}}{\delta_{\mathrm{RAO}}} \times \omega(\alpha) , \tag{24}$$

since we model the number of activated preambles as a binomial random variable of parameters NM_{D} and $\omega(\alpha)$. The service rate of the access granted phase and the number of servers are

$$\mu_D = \frac{1}{T_{\mathrm{TTI}}} \quad \text{and} \quad m_D = M_{\mathrm{D}} , \tag{25}$$

respectively. The maximum waiting time of a SR is

$$\tau_q = W \times T_{\mathrm{TTI}} . \tag{26}$$

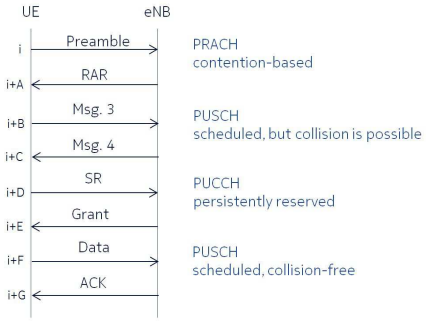

Fig. 8. LTE radio access protocol. RRC connection setup is neglected.

Finally, the access granted dropping probability is evaluated as

$$p_d^A = p_d(\lambda_A, m_D \mu_D, \tau_q)\,. \tag{27}$$

The impact of multiple transmission attempts can be evaluated as done in the One-Stage scheme, using formulas (15), (16), and (18) to evaluate the outage probability, the average number of transmissions, and the aggregate arrival rate λ_T, respectively. The system throughput can be computed as done in Eq. (19).

C. Model of the Two-Stage Protocol with Grouped Resources

In the case of grouped resources, the access granted phase is characterized as a system with K parallel queues. We recall that, for the sake of simplicity, we assume that the groups have exactly the same number of dedicated resources. The rate of activated preambles at the generic queue k is

$$\lambda_A^{[k]} = \frac{N M_{\mathrm{D}}/K}{\delta_{\mathrm{RAO}}} \times \omega(\alpha) = \frac{1}{K} \times \lambda_A \qquad \forall k = 1, \ldots, K\,. \tag{28}$$

The number of servers of a single queue becomes

$$m_D^{[k]} = \frac{M_{\mathrm{D}}}{K} = \frac{1}{K} \times m_D \qquad \forall k = 1, \ldots, K\,, \tag{29}$$

while the service rate μ_{D} remains unchanged. The dropping probability at the access grant phase is obtained using (27), but considering now the expressions of $\lambda_A^{[k]}$ and $m^{[k]}$ in place of λ_A and m_D, respectively.

D. Model of LTE for Small Packet Traffic

To provide a quantitative comparison between the proposed protocols for 5G and the current cellular standard, the LTE RA procedure has been tailored to small packet traffic and modeled following the steps shown in Fig. 8. Initially, each UE is in `RRC_IDLE` to minimize the energy consumption between two subsequent packet transmissions: indeed, the interarrival time of packets is typically much longer than the `RRC_CONNECTION_RELEASE` timer. Therefore, the UE has to switch to `RRC_CONNECTED` state through the LTE RA procedure explained in Section II. For the sake of a fair comparison, we neglect the aspects related to the RRC connection setup, in order to focus only on the core of the LTE radio access procedure.

Let us assume that n_{RB} RBs are available for PRACH, PUCCH, and PUSCH. For the sake of a fair comparison, we assume that every MTD requests a grant of fixed size[6] of J RBs, i.e., a data SPB, to send UL data on PUSCH. In the following, we describe the setup of the various physical channels and derive the model of LTE radio access procedure.

a) PRACH: The PRACH takes $n_{\mathrm{PRACH}} = 6$ RBs, stacked in frequency [8]. We set δ_{RAO} to one TTI, thus the PRACH is instantiated in every subframe. We recall that, in the entire pool of 64 Zadoff-Chu orthogonal sequences, $d = 54$ signatures are used for contention-based access and the remaining ten for contention-free access. Therefore, the number of distinguishable preambles per PRACH RB per subframe is $R_{\mathrm{PRACH}} = d/n_{\mathrm{PRACH}} = 9$.

b) PUCCH: We denote with n_{PUCCH} the number of RBs dedicated to PUCCH. This quantity must be a multiple of two, because the PUCCH is always instantiated at the opposite sides of the UL bandwidth [8]. Since PUCCH format 1 is dedicated to SRs, the maximum number of UEs that can be accomodated on PUCCH is given by

$$N_{\mathrm{PUCCH}}^{\max} = R_{\mathrm{PUCCH}} \times n_{\mathrm{PUCCH}} \times T_{\mathrm{SR}}\,, \tag{30}$$

where R_{PUCCH} is the number of orthogonal codes distinguishable per PUCCH RB and T_{SR} is SR periodicity.

c) PUSCH: The PUSCH resources are used to accomodate both CRs and data SPBs. Since the RB is the smallest resource that can be allocated in LTE, we assume that a CR message occupies exactly one RB, thus n_{CR} RBs are allocated every subframe for CR messages. Parameter n_{CR} should be upper bounded by the maximum number of UL grants that a RAR message can carry in a TTI, i.e., $n_{\mathrm{CR}} \leq 3$, but we relax this constraint assuming that the entire DL bandwidth is dedicated to small packet traffic. The number of resources dedicated to PUSCH is $n_{\mathrm{PUSCH}} = n_{\mathrm{CR}} + M_{\mathrm{D}}^{\mathrm{LTE}} \times J$, where $M_{\mathrm{D}}^{\mathrm{LTE}}$ is the number of data SPBs for LTE. We remark that it is $M_{\mathrm{D}}^{\mathrm{LTE}} \leq M_{\mathrm{D}}$, because the resources on PUSCH must be shared between CRs and data SPBs.

The values of n_{PUCCH}, n_{PUCCH}, and n_{PUSCH} are such that

$$n_{\mathrm{PRACH}} + n_{\mathrm{PUCCH}} + n_{\mathrm{PUSCH}} = n_{\mathrm{RB}}\,. \tag{31}$$

d) LTE Radio Access Protocol Model: The preamble collision probability is

$$p_c^{\mathrm{LTE}} = \mathbb{E}_j[q(j,d)] \leq 1 - \left(1 - \frac{1}{d}\right)^{\Delta - 1}\,. \tag{32}$$

As done for the Two-Stage approach, we exploit again the theory of queues with impatient customers to model the CR step. The arrival rate, service rate, and number of servers in this case are

$$\lambda_A^{\mathrm{LTE}} = \frac{d}{\delta_{\mathrm{RAO}}} \times \omega\left(\frac{\Delta}{d}\right),\ \mu_A = \frac{1}{T_{\mathrm{TTI}}}\ ,\ \text{and}\ m_A = n_{\mathrm{CR}}\,, \tag{33}$$

respectively, while the maximum waiting time is equal to the RAR window size, i.e., $\tau_A = W_{\mathrm{RAR}} \times T_{\mathrm{TTI}}$. Therefore, the drop probability of the CR phase is

$$p_d^{\mathrm{CR}} = p_d(\lambda_A^{\mathrm{LTE}}, m_A \mu_A, \tau_A)\,. \tag{34}$$

[6]Under this assumption, the UE does not need to send its Buffer Status Report (BSR), since the dimension of the data to transmit are already fixed; thus, in the following analysis we will not allocate resources for BSR messages.

TABLE IV
SYSTEM PARAMETERS FOR THE PERFORMANCE EVALUATION.

VARIABLE	VALUE
Bandwidth	20 MHz
OFDM subcarriers	1200
Subcarrier spacing	15 kHz
T	14
S	12
n_{RB}	100
n_{PUCCH}	4
n_{PRACH}	6
R_{PRACH}	9
R_{PUCCH}	18
n_{CR}	30
$M_{\mathrm{D}}^{\mathrm{LTE}}$	15
J	4
$M = M_{\mathrm{SR}} + M_{\mathrm{D}}$	24
R	9
T_{TTI}	1 ms
δ_{RAO}	1 ms

On the other hand, the data transmission takes place only if there are enough resources available. This step can be modeled as a queue with impatient customers, as well. The arrival rate is given by the number of packets that succeeded in getting a grant for the CR, i.e.,

$$\lambda_D = \lambda_S \times \left(1 - p_d^{\mathrm{CR}}\right). \tag{35}$$

While the service rate μ_D is as defined in Eq. (25), the number of servers is $m_D^{\mathrm{LTE}} = M_{\mathrm{D}}^{\mathrm{LTE}}$. The maximum waiting time is given by the SR periodicity, i.e., $\tau_D = T_{\mathrm{SR}} \times T_{\mathrm{TTI}}$. Therefore, the drop probability of the data phase is

$$p_d^D = p_d(\lambda_D, m_D^{\mathrm{LTE}}\mu_D, \tau_D). \tag{36}$$

The one-shot failure probability of the overall RA procedure is

$$p_f^{\mathrm{LTE}} = 1 - \left(1 - p_c^{\mathrm{LTE}}\right)\left(1 - p_d^{\mathrm{CR}}\right)\left(1 - p_d^D\right) \tag{37}$$

and the outage probability is

$$p_{\mathrm{outage}}^{\mathrm{LTE}} = \left(p_f^{\mathrm{LTE}}\right)^{\Theta}. \tag{38}$$

The average number of preamble transmission attempts and the aggregate arrival rate can be computed using Eq. (16) and (18). Finally, the throughput of the overall system is defined as

$$\mathcal{S}_{\mathrm{LTE}} = \lambda \times (1 - p_{\mathrm{outage}}^{\mathrm{LTE}}). \tag{39}$$

V. PERFORMANCE EVALUATION

In this section the performance of the proposed radio access protocols to support IoT in 5G networks is evaluated and compared with the LTE RA procedure. The analytical results will be compared with the computer simulation results. Finally, a discussion of the results is provided.

TABLE V
PROTOCOL PARAMETERS FOR THE PERFORMANCE EVALUATION.

VARIABLE	VALUE
Θ	4
A	0 (1-stage) 3 (2-stage) 3 (LTE)
B	0 (1-stage) $A + 1 = 4$ (2-stage) $A + 6 = 9$ (LTE)
C	3 (1-stage) $B + 3 = 7$ (2-stage) $B + 8 = 17$ (LTE)
D	$C + 4 = 21$
E	$D + 4 = 25$
F	$E + 4 = 29$
G	$F + 4 = 33$
β_{min}	0 ms
β_{max}	10 ms
T_{wake}	0.5 ms
W_{RAR}	1
$W_{\mathrm{resolution}}$	8
T_{SR}	1

A. Performance Metrics and Evaluation Assumptions

The system performance is evaluated in *ideal* conditions, i.e., assuming an error-free channel. Moreover, if two UEs in the same cell use the same resource (data or SR resource) both transmissions are lost. We will compare the three system performance metrics:

1) the *throughput*;
2) the *outage probability*;
3) and the *average transmission delay*.

In particular, the average transmission delay is defined as the period between the generation of a new data packet and the reception of final ACK. Under the assumption of independent transmission attempts, the delay $\mathcal{D}$ can be computed as in Eq. (40), where T_{TX} is the average delay of a *successful* transmission attempt, T_{RETX} is the average delay of a *unsuccessful* transmission attempt, and $\bar{\beta} = (\beta_{\mathrm{max}} - \beta_{\mathrm{min}})/2$ is the average backoff time between subsequent transmission attempts. The details about the computation of T_{TX} and T_{RETX} can be found in Appendix A.

In the following, analytical results and computer simulation results are compared considering the system parameters and the protocol parameters that can be found in Tab.s IV and V, respectively. It is worth noticing that the PHY layer parameters for LTE and 5G are the same, e.g., $R = R_{\mathrm{PRACH}}$, to provide a fair comparison. Moreover, the number of SPBs M has been obtained subtracting the RBs dedicated to PUCCH from the overall number of RBs, i.e., $M = (n_{\mathrm{RB}} - n_{\mathrm{PUCCH}})/J$.

As for the timing parameters, in LTE the processing time at the MTD side between the reception of the RAR and the CR transmission takes 5 TTIs, i.e, $B = A + 6$. In 5G, instead, the processing time at the MTD side between the reception of the ACK of the SR and the data transmission can be minimized due to the optimized feedback design described in Section III-D, thus we assume that $B = A + 1$. Moreover, a mean waiting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2707074, IEEE Transactions on Communications

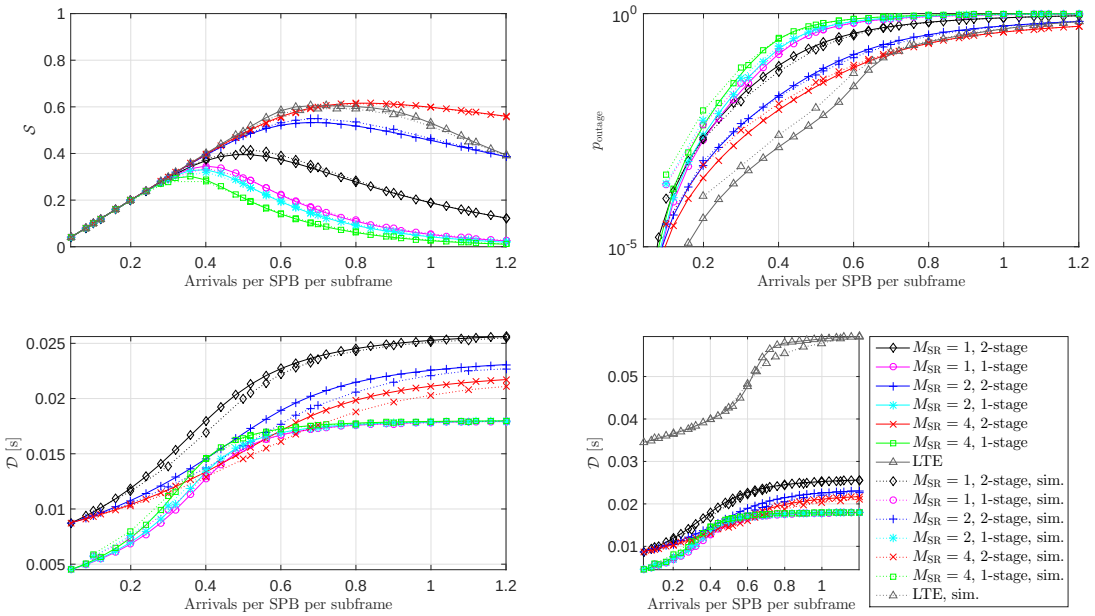

Fig. 9. Impact of over-provisioning factor N, i.e., the trade-off between SR resources M_{SR} and tagged data resources M_{D}. M_{SR} is varied in $\{1, 2, 4\}$, $K = M_{\text{D}}$ and no windowing is allowed ($W = 1$). The analytical results and the simulation results are represented in solid lines and dashed lines, respectively. LTE parameters according to Tab.s IV and V.

time between the MTD wake-up and the beginning of the next TTI T_{wake} of half TTI is considered. We neglect, instead, the possible offsets between UL and DL frame, the propagation time, as well as the time required for the wake-up process of the device and the delay introduced by an additional final ACK from the application layer, which could be quantified in a few additional milliseconds on aggregate. Also, we remark that our definition of delay includes the time between data transmission and reception of the ACK, i.e., the duration $C - B$. However, in practice, as soon as the data is successfully decoded at the eNB, it may be already forwarded to its final destination. Thus, the duration $C - B$ will not extend the overall end-to-end delay of the service.

B. Pure Protocol Performance

Four studies have been made to test the performance of the proposed protocols, varying parameter N, K, W, and R, respectively.

1) Impact of Over-Provisioning Factor N: The graphical results can be found in Fig. 9, where we assume $M_{\text{SR}} \in \{1, 2, 4\}$, tagged data resources, i.e., $K = M_{\text{D}}$, and no windowing ($W = 1$). The solid lines and dashed lines denote the results of the theoretical model and the numerical evaluation, respectively. It can be seen that the Two-Stage protocol with tagged resources outperforms the One-Stage protocol in terms of throughput, failure probability, and outage probability, while the One-Stage protocol provides a faster packet delivery of successful attempts if the arrival rate is sufficiently low. For high arrival rates, indeed, the outage probability of the One-Stage protocol approaches one, meaning that very few packets are successfully delivered. Moreover, as M_{SR} increases, the One-Stage protocol performance is degraded, while the Two-Stage protocol performance improves, as expected. Finally, we want to remark that the theoretical curves and the empirical curves nicely overlap in terms of throughput, failure probability, outage probability, and average number of transmission attempts of successful packets. The greatest difference is in the delay plots, where the gap between the theoretical evaluation and the empirical evaluation in the Two-Stage protocol is due to the assumption of statistical independence between the two stages of the transmission as well as among successive transmission attempts in the theoretical model.

The gains of 5G protocols over LTE mainly regard the delay, because of the additional signaling exchange shown in Fig. 8. Please also note that the impact of MUD is not yet included in this analysis. A further increase of throughput is expected

$$\begin{aligned}\mathcal{D} &= \sum_{\theta=1}^{\Theta}[T_{\text{TX}} + (\theta-1)\times(T_{\text{RETX}} + \bar{\beta})] \times \mathbb{P}[\theta \text{ tx attempts}|\text{final ACK}] \\ &= \sum_{\theta=1}^{\Theta}[T_{\text{TX}} + (\theta-1)\times(T_{\text{RETX}} + \bar{\beta})] \times \frac{p_f^{\theta-1}(1-p_f)}{1-p_f^{\Theta}} = T_{\text{TX}} + (T_{\text{RETX}} + \bar{\beta}) \times \frac{p_f - \Theta p_f^{\Theta} + (\Theta-1)p_f^{\Theta+1}}{(1-p_f^{\Theta})(1-p_f)}\,.\end{aligned} \tag{40}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2707074, IEEE Transactions on Communications

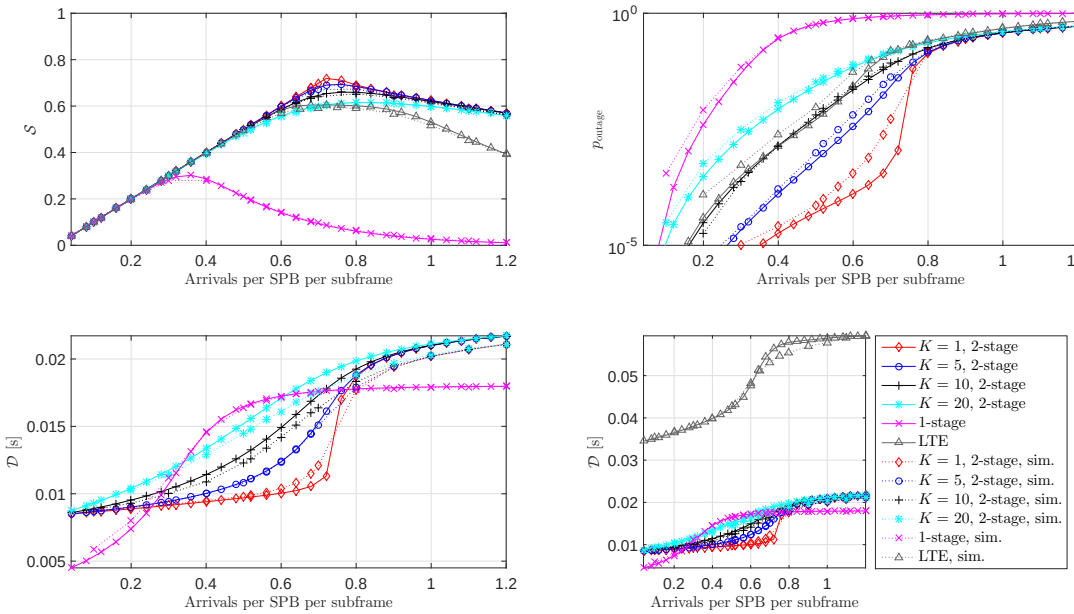

Fig. 10. Impact of grouping with parameter K for $M_{\text{SR}} = 4$ without windowing ($W = 1$). The analytical results and the simulation results are represented in solid lines and dashed lines, respectively.

if a sophisticated receiver could decode at least one or more packets in spite of a collision. This aspect will be investigated as future work.

2) Impact of the Number of Groups K: The graphical results can be found in Fig. 10, where we fix $M_{\text{SR}} = 4$, $W = 1$, and let K vary in the set $\{1, 5, 10, 20\}$. We observe that the increase in the number of groups K degrades the performance of the Two-Stage protocol; therefore, the pooled version is more efficient than the tagged version due to the enhanced flexibility for the assignment of data SPBs at the eNB. However, as already discussed, the benefit of the tagged variant is the smaller DL feedback size. Moreover, grouping may be needed for efficient service differentiation and prioritization.

3) Impact of Window Size W: The graphical results can be found in Fig. 11, where it is $M_{\text{SR}} = 4$, $K = M_{\text{D}} = 20$, and $W \in \{1, 2, 4, 8\}$. The time flexibility results to be beneficial if associated with tagged data SPBs. Indeed, it can be seen that an increase in the window size W boosts the performance of the Two-Stage protocol. The time window W does not provide an additional benefit if the pooled variant of the Two-Stage protocol is applied. The reason is that the potential of increased flexibility is already fully exploited in frequency direction as explained above. Thus, tagged resources combined with time windowing (see Fig. 11, $W = 8$) can be seen as equivalent solution to pooled resources (see Fig. 10, $K = 1$).

4) Impact of PHY: Finally, we investigate the impact of the number of detectable preamble sequences per RB R. An increase of R is equivalent to a higher over-provisioning factor N, however without the need to reserve a larger portion of the radio resources for SRs, i.e., in contrast to Fig. 9 we keep the values for M_{SR} and M_{D} constant. With a novel preamble sequence design like the one introduced in [41] at least a duplication of the number of preamble sequences can be achieved, i.e., $R = 18$ instead of $R = 9$. The comparison is shown in Fig. 12. Obviously, all performance metrics clearly benefit from the higher number of preambles due to a significantly smaller collision probability. The dynamic increase of the over-provisioning factor N by switching from $R = 9$ to $R = 18$ is an important means to mitigate a congestion through sporadic massive access of arrivals. The drawback, however, is a slightly increased probability for missed detections and false alarms of SRs. Please remind that we assume perfect preamble detection capabilities throughout this paper.

VI. Conclusion

We have presented two efficient radio access protocols for sporadic small UL data traffic aiming at minimal signaling overhead and scalability with respect to the number of IoT devices per radio cell. Their main characteristic is to transmit the data already in the first or in the second stage of the communication. In contrast to LTE, the two protocols are *contention-based* and eventually *connectionless*, i.e., there is no collision resolution mechanism and connection setup and release are not required before and after data transmission. The main advantage of the One-Stage variant of the protocol is its minimal delay in case of low traffic load. On the other hand, the Two-Stage variant of the protocol is superior with respect to throughput and is the appropriate choice at high traffic load. The latter solution allows building pools of SPBs in order to provide flexibility for the assignment of resources, according to the QoS of single MTDs. Moreover, a compensation for single overloaded TTIs can be realized by introducing a time window for data transmission. Both improvements come along with a slightly increased number of required DL feedback bits. This drawback can be mitigated with the proposed simplified

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2707074, IEEE Transactions on Communications

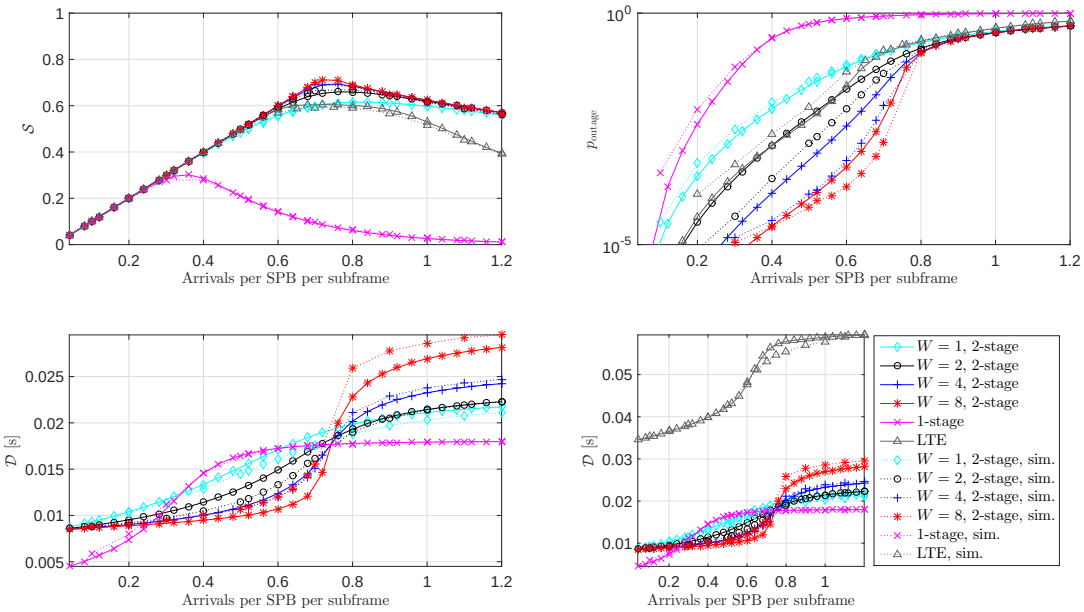

Fig. 11. Impact of windowing with parameter W for $M_{\text{SR}} = 4$ and $K = M_{\text{D}}$. The analytical results and the simulation results are represented in solid lines and dashed lines, respectively.

queueing feedback scheme, which is our envisaged solution for the 5G air interface design.

A mathematical model has been derived to analytically evaluate the performance of the proposed protocol variants with respect to throughput, outage probability, and delay. The model has been validated by system-level simulation. Firstly, we have investigated the trade-off between SR and data resources: a high over-provisioning of SR opportunities reduces the collision probability, but the limitation of SPBs for data packets impacts the throughput. Secondly, we have analyzed the impact of grouping of data SPBs: with pooled resources we can achieve the best performance for all metrics at the cost of a higher number of feedback bits. Finally, we showed that a similar effect can be in principle achieved with windowing, at the cost of an additional delay.

As for future work on this topic, we will exploit MUD techniques in the protocol design and investigate their impact on the system performance.

APPENDIX A
COMPUTATION OF T_{TX} AND T_{RETX}

In this section the computation of the average delay of a successful transmission attempt, denoted with T_{TX}, and of a failed transmission attempt, denoted with T_{RETX}, is provided.

One-Stage Protocol

In the One-Stage protocol T_{TX} is simply given by

$$\begin{aligned} T_{\text{TX}} &= T_{\text{wake}} + [(i + C) - i + 1] \times T_{\text{TTI}} \\ &= T_{\text{wake}} + (C + 1) \times T_{\text{TTI}} \end{aligned} \tag{41}$$

and T_{RETX} is equal to the average transmission time without the wake-up time, i.e.,

$$T_{\text{RETX}} = T_{\text{TX}} - T_{\text{wake}} = (C + 1) \times T_{\text{TTI}} \,. \tag{42}$$

Two-Stage Protocol

In the Two-Stage protocol without windowing, i.e., $W = 1$, T_{TX} is expressed as in Eq. (41). For window sizes W such that $W > 1$, instead, we must account for the average delay introduced by the window W. This can be done exploiting the theory of queues with impatient customers [42]. The relationship between queue dropping probability p_d, worst case average wait time τ, and average waiting time $\bar{W}_{\text{wait}}$ is defined as

$$p_d = \frac{\bar{W}_{\text{wait}}}{\tau} \,, \tag{43}$$

thus the average waiting time is computed as

$$\bar{W}_{\text{wait}} = p_d \times \tau \,. \tag{44}$$

In the case of the Two-Stage approach, we have to plug in the values of p_d^A in Eq. (27) and τ_q.

The successful transmission interval duration, then, can be derived as

$$T_{\text{TX}} = T_{\text{wake}} + (C + \bar{W}_{\text{wait}} + 1) \times T_{\text{TTI}} \,, \tag{45}$$

while T_{RETX} is obtained averaging between the delays introduced if a failure occurs after the preamble transmission or after the data transmission, i.e.,

$$T_{\text{RETX}} = p_d^A \times (B + 1) \times T_{\text{TTI}} + \left(1 - p_d^A\right) \times (T_{\text{TX}} - T_{\text{wake}}) \,. \tag{46}$$

LTE

In order to make the packet transmission as fast as possible, we set the LTE protocol parameters to their minimum values, i.e., $W_{\text{RAR}} = 1$, $W_{\text{resolution}} = 8$, and $T_{\text{SR}} = 1$, as stated in Tab. V.

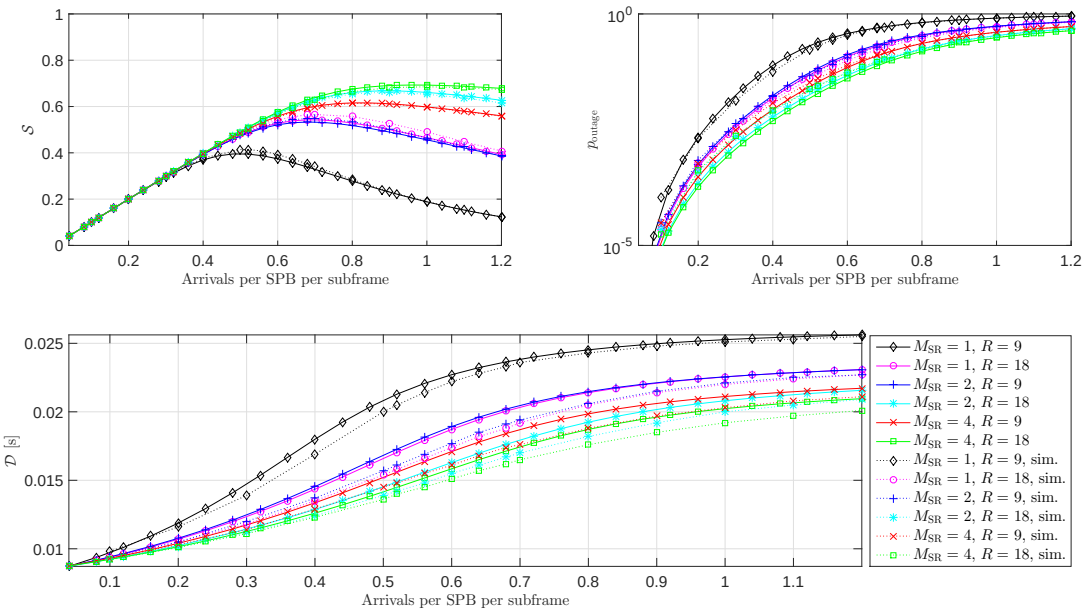

Fig. 12. Impact of PHY design, according to parameter R, for $M_{SR} = \{1, 2, 4\}$, $K = M_D$, and $W = 1$.

As a consequence, no time flexibility is allowed. The successful transmission attempt duration is

$$T_{TX} = T_{wake} + (G + 1) \times T_{TTI}, \quad (47)$$

while the retransmission time is

$$T_{RETX} = [p_d^A \times (A + 1) + (1 - p_d^A) \times (C + 1)] \times T_{TTI}. \quad (48)$$

ACKNOWLEDGMENT

Part of this work has been performed in the framework of the Horizon 2020 project FANTASTIC-5G (ICT-671660), which is partly funded by the European Union: the authors would like to acknowledge the contributions of their colleagues in FANTASTIC-5G. The authors would like to thank also the Editor and the anonymous reviewers for their valuable comments.

REFERENCES

[1] S. Saur, A. Weber, and G. Schreiber, "Radio access protocols and preamble design for machine type communications in 5G," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2015, pp. 8–12.

[2] M. Centenaro and L. Vangelista, "Analysis of Small Packet Traffic Support in LTE," in *Proc. Wireless Telecommunications Symp. (WTS)*, Chicago, USA, Apr. 2017.

[3] NGMN, "5G White Paper," Feb. 2015. [Online]. Available: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf

[4] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," Tech. Rep., Feb. 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf

[5] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, 2015, Accepted Manuscript. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S235286481500005X

[6] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[7] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*. John Wiley & Sons, Inc., 2007.

[8] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.

[9] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Inc., 2009.

[10] 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 Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, First Quarter 2014.

[11] 3GPP, "Study on RAN Improvements for Machine-type Communications," Tech. Rep. TR 37.868 V11.0.0, Sep. 2011.

[12] ——, "Study on Enhancements to Machine-Type Communications (MTC) and other Mobile Data Applications," Tech. Rep. TR 37.869 V12.0.0, Sep. 2013.

[13] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. P. Koskinen, "Overview of narrowband IoT in LTE Rel-13," in *Proc. IEEE Conf. Standards for Communications and Networking (CSCN)*, Berlin, Germany, Oct. 2016, pp. 1–7.

[14] C.-Y. Tu, C.-Y. Ho, and C.-Y. Huang, "Energy-Efficient Algorithms and Evaluations for Massive Access Management in Cellular Based Machine to Machine Communications," in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, San Francisco, CA, USA, Sep. 2011, pp. 1–5.

[15] S.-Y. Lien and K.-C. Chen, "Massive Access Management for QoS Guarantees in 3GPP Machine-to-Machine Communications," *IEEE Commun. Lett.*, vol. 15, no. 3, pp. 311–313, Mar. 2011.

[16] S. Bayat, Y. Li, Z. Han, M. Dohler, and B. Vucetic, "Distributed massive wireless access for cellular machine-to-machine communication," in *Proc. IEEE Int. Conf. Communications (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 2767–2772.

[17] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[18] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, "Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine (M2M) Communications," in *Proc. World Wireless Research Forum Meeting (WWRF)*, Düsseldorf, Germany, Oct. 2011.

[19] Y. Liu, C. Yuen, X. Cao, N. U. Hassan, and J. Chen, "Design of a Scalable Hybrid MAC Protocol for Heterogeneous M2M Networks," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 99–111, Feb. 2014.

[20] E. Paolini, C. Stefanovic, G. Liva, and P. Popovski, "Coded random access: applying codes on graphs to design random access protocols," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144–150, Jun. 2015.

[21] F. Monsees, M. Woltering, C. Bockelmann, and A. Dekorsy, "Compressive Sensing Multi-User Detection for Multicarrier Systems in Sporadic Machine Type Communication," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Glasgow, UK, May 2015, pp. 1–5.

[22] A. Ijaz, L. Zhang, M. Grau, A. Mohamed, S. Vural, A. U. Quddus, M. A. Imran, C. H. Foh, and R. Tafazolli, "Enabling Massive IoT in 5G and Beyond Systems: PHY Radio Frame Design Considerations," *IEEE Access*, vol. 4, pp. 3322–3339, Jun. 2016.

[23] F. Schaich, T. Wild, and R. Ahmed, "Subcarrier Spacing - How to Make Use of This Degree of Freedom," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Nanjing, China, May 2016, pp. 1–6.

[24] M. Fuhrwerk, J. Peissig, and M. Schellmann, "On the design of an FBMC based AIR interface enabling channel adaptive pulse shaping per sub-band," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Nice, France, Aug. 2015, pp. 384–388.

[25] L. Zhang, A. Ijaz, P. Xiao, A. Quddus, and R. Tafazolli, "Subband Filtered Multi-carrier Systems for Multi-service Wireless Communications," *IEEE Trans. Wireless Commun.*, Jan. 2017, preprint, DOI:10.1109/TWC.2017.2656904.

[26] W. Xu and G. Campbell, "A near perfect stable random access protocol for a broadcast channel," in *Proc. IEEE Int. Conf. Communications (ICC)*, Chicago, IL, USA, Jun. 1992, pp. 370–374, vol. 1.

[27] J. Alonso-Zarate, C. Verikoukis, E. Kartsakli, A. Cateura, and L. Alonso, "A near-optimum cross-layered distributed queuing protocol for wireless LAN," *IEEE Wireless Commun.*, vol. 15, no. 1, pp. 48–55, Feb. 2008.

[28] 3GPP, "Medium Access Control (MAC) protocol specification," Tech. Rep. TS 36.321 V9.0.0, Oct. 2009.

[29] ——, "Multiplexing and channel coding," Tech. Rep. TS 36.212 V8.8.0, Jan. 2010.

[30] C. Kahn and H. Viswanathan, "Connectionless access for mobile cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 26–31, Sept. 2015.

[31] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area Wireless Communication Challenges for the Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 168–174, Feb. 2017.

[32] K. Zhou, N. Nikaein, R. Knopp, and C. Bonnet, "Contention Based Access for Machine-Type Communications over LTE," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Yokohama, Japan, May 2012, pp. 1–5.

[33] R. P. Jover and I. Murynets, "Connection-less communication of IoT devices over LTE mobile networks," in *Proc. IEEE Int. Conf. Sensing, Communication, and Networking (SECON)*, Seattle, WA, USA, Jun. 2015, pp. 247–255.

[34] F. Schaich and T. Wild, "Relaxed synchronization support of universal filtered multi-carrier including autonomous timing advance," in *Proc. Int. Symp. Wireless Communications Systems (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 203–208.

[35] L. Zhang, P. Xiao, A. Zafar, A. ul Quddus, and R. Tafazolli, "FBMC System: An Insight into Doubly Dispersive Channel Impact," *IEEE Trans. Veh. Technol.*, Aug. 2016, preprint, DOI:10.1109/TVT.2016.2602096.

[36] X. Zhang, M. Jia, L. Chen, J. Ma, and J. Qiu, "Filtered-OFDM - Enabler for Flexible Waveform in the 5th Generation Cellular Networks," in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[37] J. J. Nielsen, D. Kim, G. C. Madueño, N. K. Pratas, and P. Popovski, "A Tractable Model of the LTE Access Reservation Procedure for Machine-Type Communications," in *Proc. IEEE Global Communication Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[38] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[39] A. G. D. Kok and H. G. Tijms, "A Queueing System with Impatient Customers," *Journal of Applied Probability*, vol. 22, no. 3, pp. 688–696, Sept. 1985. [Online]. Available: http://www.jstor.org/stable/3213871

[40] N. K. Boots and H. Tijms, "A multiserver queueing system with impatient customers," *Management Science*, vol. 45, no. 3, pp. 444–448, Mar. 1999.

[41] J. C. Guey, "The Design and Detection of Signature Sequences in Time-Frequency Selective Channel," in *Proc. Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, Sep. 2008, pp. 1–5.

[42] S. Zeltyn, "Call centers with impatient customers: exact analysis and many-server asymptotics of the M/M/n+G queue," Ph.D. dissertation, Israel Institute of Technology, Oct. 2004.

Marco Centenaro (S'14) received the Bachelor's degree in Information Engineering in 2012 and the Master's degree in Telecommunication Engineering in 2014, both from the University of Padova, Italy. Since November 2014 he is a Ph.D. student at the Department of Information Engineering of the University of Padova, Italy. From September 2016 to December 2016 he was a visiting student at Nokia Bell Labs, Stuttgart, Germany. His research interests include the next generation of cellular networks (5G) and in particular the Machine-to-Machine Communication.

Lorenzo Vangelista (S'93-M'97-SM'02) received the Laurea and Ph.D. degrees in electrical and telecommunication engineering from the University of Padova, Padova, Italy, in 1992 and 1995, respectively. He subsequently joined the Transmission and Optical Technology Department, CSELT, Torino, Italy. From December 1996 to January 2002, he was with Telit Mobile Terminals, Trieste, Italy, and then, until May 2003, he was with Microcell A/S, Copenaghen, Denmark. In July 2006, he joined the Worldwide Organization of Infineon Technologies as Program Manager. Since October 2006, he has been an Associate Professor of Telecommunication with the Department of Information Engineering, Padova University. His research interests include signal theory, multicarrier modulation techniques, cellular networks and wireless sensors and actuators networks.

Stephan Saur (M'08) received the Dipl.-Ing. and Dr.-Ing. degrees in Electrical Engineering from the University of Stuttgart in 2000 and 2008, respectively. He joined Alcatel Research & Innovation in 2006 and is now member of technical staff in Nokia Bell Labs Wireless Research, where he is entrusted with the development of novel PHY, MAC, and system concepts for future cellular mobile radio systems. His current research topic is the energy- and spectral efficient radio access for sporadic low-rate traffic.

Andreas Weber is team leader of the mobile system performance evaluation group in Bell Labs' Wireless Domain in Stuttgart, Germany. He received Dipl.-Ing. and Dr.-Ing. Degrees in electrical engineering from the University of Stuttgart, Germany. Prior to joining Nokia, Dr. Weber worked in the field of satellite communications as a member of scientific staff at the Institute of Communications Switching and Data Technics, University of Stuttgart. During his tenure at Alcatel Research & Innovation and later at Bell Labs, he worked on the performance evaluation and optimization of 2G, 3G, 4G, and beyond 4G mobile communication systems. Currently, he works on 5G system performance evaluation, especially with respect to uplink small packet access, downlink control channel performance, and channel quality feedback.

Volker Braun is a project leader in the wireless access domain at Nokia Bell Labs, Stuttgart, Germany. He led the development of the base station software implementing the HSPA fast radio resource management, as well as the definition and integration of an LTE pre-standard prototype system used for early customer mobility trials. Currently, he is leading research work related with 5G air interface and system design, both low band and high band. Dr. Braun contributed to 3GPP RAN and RAN1 working groups, to the LTE/SAE Trial Initiative and to funded research projects (EASY-C, ARTIST4G, METIS). He holds a doctoral degree in engineering.