

Group Paging-Based Energy Saving for Massive MTC Accesses in LTE and Beyond Networks

Osama Arouk, *Student Member, IEEE*, Adlen Ksentini, *Senior Member, IEEE*, and Tarik Taleb, *Senior Member, IEEE*

Abstract—Next generation cellular networks (5G) have to deal with massive deployment of machine-type-communication (MTC) devices, expected to cause congestion and system overload in both the radio access network (RAN) and the core network (CN). Moreover, not only would the network suffer from the system overload, but also the MTC devices would experience high latency to access the channel and high power consumption due to the retransmission attempts. Indeed, power consumption is a critical issue in MTC, as the devices are not plugged into the electrical supply, e.g., in the case of sensor devices. To alleviate system overload (caused by the massive MTC deployment), the 3GPP proposed the group paging (GP) method. However, its performances dramatically decrease when increasing the number of MTC devices being paged. In this paper, we devise a novel method, named further improvement-traffic scattering for group paging (FI-TSFGP), which aims to improve the performance of GP when the number of MTC devices is high. FI-TSFGP scatters the paging operation of the MTC devices over a GP interval instead of letting all of the devices start the channel access procedure at nearly the same time. By doing so, FI-TSFGP achieves high-channel access probability for MTC devices, leading to the reduction of both the channel access latency and power consumption. Compared to GP and two other schemes, simulation results clearly demonstrate the high performance of FI-TSFGP in terms of: success and collision probabilities, average access delay, average number of preamble transmissions, and ultimately energy conservation.

Index Terms—MTC, M2M, massive MTC, Energy efficient, LTE, LTE-A, 5G, congestion control, overload control, RACH procedure, group paging.

I. INTRODUCTION

ONE OF THE main 5G requirements is to ensure the connection of massive numbers of wireless devices to cellular networks, including not only the User Equipments (UEs) but also objects like sensors and actuators that constitute the concept of Internet of Things (IoT) or Machine-To-Machine Communications (Machine Type Communications) [1]. According to its definition, MTC can be viewed as an emerging technology referring to the communication between machines (devices) without (or with a little) human intervention. Under

Manuscript received March 29, 2015; revised October 19, 2015; accepted December 6, 2015. Date of publication January 21, 2016; date of current version May 19, 2016. This work was supported by the TAKE 5 project funded by the Finnish Funding Agency for Technology and Innovation (TEKES), a part of the Finnish Ministry of Employment and the Economy.

O. Arouk and A. Ksentini are with the IRISA Laboratory, University of Rennes 1, Cedex 35042, France (e-mail: osama.arouk@gmail.com; adlen.ksentini@irisa.fr).

T. Taleb is with Sejong University and Aalto University, Espoo FI-00076, Finland (e-mail: talebtarik@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSAC.2016.2520222

the aforementioned vision, MTC would support a large number of applications, in various domains, such as Healthcare (eHealth), Intelligent Transport System (ITS), smart grid and smart metering, Public Safety (PS), etc.

Energy conservation represents an important factor for the successful deployment of MTC devices. Particularly, for massive MTC deployment case (noted by massive MTC), whereby the devices need to be low-cost, affordable and able to operate on a battery for several years, i.e. strict requirements for low power consumption (e.g., 10 times longer battery life). Indeed, recent forecasts predict that: (i) there would be 50 billion MTC devices by 2020 [2]; (ii) MTC traffic would increase 24 times by 2017 compared to 2012; (iii) the total traffic volume, in the wireless communication systems, would be increased 1000 times compared to today's traffic volume [3]–[5]. Massive MTC would not only impact the cellular network functioning by introducing system overload and congestion, but also the MTC devices in terms of energy consumption. Indeed, massive MTC will generate a huge amount of data/control traffic, leading to congestion and system overload in both the RAN and CN parts. This congestion may cause intolerable delay, packet loss, or even service unavailability for both MTC and Non-MTC traffic. At the same time, MTC devices will experience low success in accessing the channel access, thus increasing the retransmission attempts and dramatically increasing the energy consumption. In this context, it is important to devise mechanisms that alleviate system overload and consequently increase the MTC devices' battery lifetime by reducing energy consumption.

Group Paging (GP) is an effective solution proposed by the 3GPP group to alleviate the congestion in 4G networks. In the GP method, the MTC devices are grouped together according to various metrics, such as time-controlled, delay-tolerant, Quality of Service (QoS), etc. Each group is assigned an ID, named Group ID (GID). When the network needs some information from a certain group, it sends a paging message addressed by its ID, i.e. GID. Once receiving the paging message, all the members of this group will start the contention-based Random Access CHannel (RACH) procedure in the first available Random Access (RA) resources [6]–[8]. In spite of its advantages, the performance of GP method dramatically decreases when increasing the number of MTC devices being paged. An improvement of the GP, namely Controlled Distribution of Resources (CDR), has been introduced in [9], whereby a scheduling based on terminal ID in the cell, i.e. Cell-Radio Network Temporary Identifier (C-RNTI), is used. Whilst the CDR method highly improves the performance compared to GP method, it is only dedicated to the case when

MTC devices are in the RRC_CONNECTED mode, ignoring the RRC_IDLE mode. The authors in [10] propose to repeat the group paging interval, i.e. Consecutive Group Paging (CGP), so that the MTC devices having not succeeded in the first GP interval will try to access the network in the subsequent GP interval(s). However, CGP performances are worse than the classical Group Paging [6] for certain configuration. In [11], and similar in spirit to the idea of [12], the authors proposed enforcing some backoff time on new transmission attempts before the first preamble transmission, i.e. pre-backoff. Results presented in [11] showed the superiority of pre-backoff (PBO) method by report to the classical GP method. Another improvement of the GP method was introduced in [13], wherein a new method is devised, namely Traffic Scattering For Group Paging (TSFGP), which highly improves the performance, compared to GP, regardless the state of the device. In this paper we introduce a Further Improved version of TSFGP (i.e., dubbed as FI-TSFGP) that enhances TSFGP performance leveraging a better estimation of both the total number of arrivals and the number of successful MTC devices in the stable state. FI-TSFGP accurately estimates the latter, whatever the network's parameters (e.g., the number of preamble transmissions $N_{PT_{max}}$ and the number of available preambles R), which is not the case of TSFGP that fails for certain configurations (e.g., when $N_{PT_{max}}$ is large and R is small). Accordingly, FI-TSFGP gives the flexibility to change the network's parameters, e.g. changing the GP interval by changing the number of preamble transmissions. On the other hand, FI-TSFGP shares the same objective as PBO, i.e. reducing the collisions during the RACH procedure. To achieve this objective, PBO spreads the MTC devices (via pre-backoff operation) over a certain interval regardless their number, while FI-TSFGP activates the number of MTC devices that maximizes the performances (e.g., maximizing the success probability and the resource utilization). Thanks to this difference (as illustrated in section V), FI-TSFGP maintains good performances whatever the number of MTC devices, while PBO performances degrade when the number of MTC devices increases. Compared to GP [6], CGP [10], and PBO [11], FI-TSFGP highly outperforms these methods in terms of success and collision probabilities, average access delay, average number of preamble transmissions, and ultimately energy efficiency.

The remainder of the work is organized as follows. Section II introduces a quick overview of the related works. Some background about MTC system architecture, and RACH procedure is introduced in section III. In section IV, system model, used in our study, and the analysis of our proposition are detailed. The performance results of FI-TSFGP, GP, CGP, and PBO are presented and compared in section V. Finally, conclusions are presented in section VI.

II. RELATED WORK

Congestion and system overload, that may occur when deploying MTC in LTE, are usually tackled by using different techniques, such as increasing the available resources [14] or by throttling/controlling the traffic. Based on which entity (i.e., UE or evolved Node B - eNB) initiates the Random Access

CHannel (RACH) procedure, existing solutions can be classified into two categories: Push and Pull based approaches. In the push category, the RACH procedure is initiated by the terminals (UEs or MTC devices), which yields to consider it also as a decentralized control scheme. Many methods fall in this category [15]:

1. Separate RACH resources: when sharing the resources between M2M and Human-to-Human (H2H), there will be a large impact on the QoS of H2H as the number of MTC is naturally larger. Separating the RACH resources between M2M and H2H is a requirement to limit the impact of MTC on H2H traffic. This separation could be done through different ways; time, frequency, preamble separations, or a mixture of them. However, the disadvantage of this scheme is that the resources of one type, for example MTC's dedicated resources, can not be used by another type having more traffic, even when the first type does not have traffic to send. Based on the latest observation, static separation of resources is not a good option. Another solution consists in separating the available resources into two groups: the first one is dedicated to H2H, and the another one is shared between H2H and M2M [16]. Note that simulation results in [16] have proved that this method of separation outperforms the static separation.
2. Dynamic allocation of RACH resources: this scheme can be viewed as an improvement of the precedent one, since the resources are dynamically allocated based on the predicted traffic. Though this scheme better handles the congestion's problem, it can be used only when the network is aware about the time when the MTC devices have information to be sent.
3. Access Class Barring (ACB) Scheme: by introducing a separate access class(es) for MTC devices, ACB allows the network to control the access of MTC devices separately, avoiding any impact or penalty on the Non-MTC traffic. The granularity of the Access Class could be extended to even distinguish between MTC classes, i.e. to introduce priority between MTC applications. When the ACB method is used, the network broadcasts two parameters: (i) *acb_BarringFactor* that represents the probability of barring; (ii) *acb_BarringTime* that determines the duration in which the terminal should back off, before retrying the RACH procedure (if it fails to pass the ACB check). In the literature, there are many methods targeting the dynamic changes of the ACB parameters, especially the parameter *acb_BarringFactor*, such as [17] which tries to adjust *acb_BarringFactor* using the Proportional Integrative Derivative (PID) controller, and [18] where a traffic prediction is used in order to adjust the *acb_BarringFactor* in the case of Beta traffic as specified by 3GPP [15]. The authors in [19] introduced a new RACH procedure engineered for M2M communication, which is essentially a combination of the conventional RACH procedure with ACB. The advantage of this scheme consists in the fact that it allows the MTC devices to transmit their data just after the preamble transmission. It also has a self-optimization feature, allowing the system to achieve optimal MTC throughput.

Regarding the second category, i.e. pull based scheme, the network (eNB) initiates the RACH procedure. This category is also known as centralized control solutions. There are several Pull based congestion control methods, among them we can cite the Paging and Group Paging (GP) methods. In the Paging method, the network sends a paging message when it needs some information from a certain terminal. This method is a rational one when paging a few number of devices, while it becomes impractical when paging a large number of MTC devices. As an example, paging 36000 MTC devices will require about 11.5 s if we know that there are two paging occasions in each radio frame (10 ms) and at most 16 MTC devices can be paged by each paging occasion. One solution of this issue is to use the GP method, whereby all the members of the group are paged by just one paging message, addressed by GID [6], [7].

Given that Push based approaches are decentralized control solutions, the resource utilization would not be stable, and it might be degraded in the presence of a large number of devices. Besides, it is so difficult to regulate the network load as the traffic is originated by the devices. However, there are some advantages of using this category. For example, the signaling load will be low, as there is no need for paging messages. Further, this category is adequate for unscheduled events, such as detection of the fire in the forest. On the other hand, in the Pull based approaches, the resource utilization would be more stable due to the fact that the control is totally held by the network. Furthermore, the network load would be easily regulated. The disadvantage of this category is the signaling load, which will be slightly higher because of the paging message(s). Pull based approach is also inadequate for unscheduled events. Indeed, when it is applied with unscheduled events, the network has to send every time a paging message to know whether the devices need to send information or not. Pull based approach will be costly if it is applied with unscheduled events. In spite of its disadvantages, it is preferable, from the network's viewpoint, to use the Pull based approach, rather than Push based one, whenever it is possible as the control is totally held by the network.

III. MTC IN THE 4G LANDSCAPE

A. MTC Network Architecture

Fig. 1 illustrates the envisioned 3GPP architecture to support MTC [20]. It consists of three main domains: the MTC domain, the communication network domain, and the MTC application domain. The MTC application domain comprises MTC servers, which are under the control of the mobile network operator or a third party. Two new entities relative to MTC communication have been recently added to the 3GPP architecture: the MTC InterWorking Function (MTC-IWF) and the Services Capability Server (SCS). SCS is an entity connecting MTC application servers to the 3GPP network so as to enable them to communicate through specific services, defined by 3GPP, with MTC and MTC-IWF. The SCS can be connected to one or more MTC-IWFs and it is controlled by the operator of the Home Public Land Mobile Network (HPLMN) or by a

third party [21]. On the other hand, MTC-IWF hides the internal topology of the Public Land Mobile Network (PLMN) and relays or translates signaling protocols used over Tsp (a reference point used by a SCS to communicate with the MTC-IWF related control plane signaling) in order to invoke specific functionality inside the PLMN. There are one or more instances of MTC-IWF in the HPLMN and it can be a standalone entity or a functional entity of another network element, with the ability to connect to one or more SCSs [20]. As shown in Fig. 1, there are three ways for establishing communication between MTC servers and MTC devices: direct model, indirect model, and hybrid model [20]. In the direct model, a MTC server is directly connected to the operator's network in order to perform user plane communications with the devices (UE or MTC) without using any SCS. In the indirect model, the MTC server indirectly connects through the services of a SCS to the operator's network. The hybrid model is when the direct and indirect models are used simultaneously.

B. RACH Procedure

A terminal trying to connect to the network must perform Radio Resource Control (RRC) connection setup procedure (see Fig. 2) [22], [23]. The first four signaling steps concern the random access procedure, also known as Initial Ranging (IR) [24]–[27], and they are detailed below. It should be noted that lots of research work is being conducted to accommodate the RACH procedure with millimeter Waves (mmWave) beamforming cellular networks, which is expected to be used in 5G [28], [29]. In general, there are two forms of random access procedure: contention-based and contention-free random access procedures. The first one is used, for example, when a terminal is moving from RRC_IDLE to RRC_CONNECTED, or trying to recover the uplink synchronization, while the second one is used, for example, for handover or DownLink (DL) data arrival [23]. The steps of the RACH procedure are as follow (also depicted in Fig. 2):

1) *Random Access Preamble Transmission (Msg1)*: The first step consists in transmitting a randomly chosen preamble. This step allows the eNB to estimate the transmission timing of the terminal that would be later used for adjusting the uplink synchronization. The frequential temporal resource in which the preamble is transmitted is known as the Physical Random Access CHannel (PRACH). As the preamble is randomly chosen, we may have the case that more than one terminal choose the same preamble, thus causing a collision. Another important objective of this step is to adjust the power transmission of the terminal, which is achieved by the power ramping factor that is Power Ramping Step (*PRS*) in equation (2). For the first time of preamble transmission, all the terminals in the cell will transmit with the same power. The received power level of the signals transmitted by terminals close to the base station, i.e. eNB, would be enough to be detected, while this level for those far from the eNB may not be sufficient to be detected. In the latter situation, these terminals will retransmit the preamble with a power level *PRS* dB higher than the one used in the precedent attempt. The advantage of this technique is that each terminal

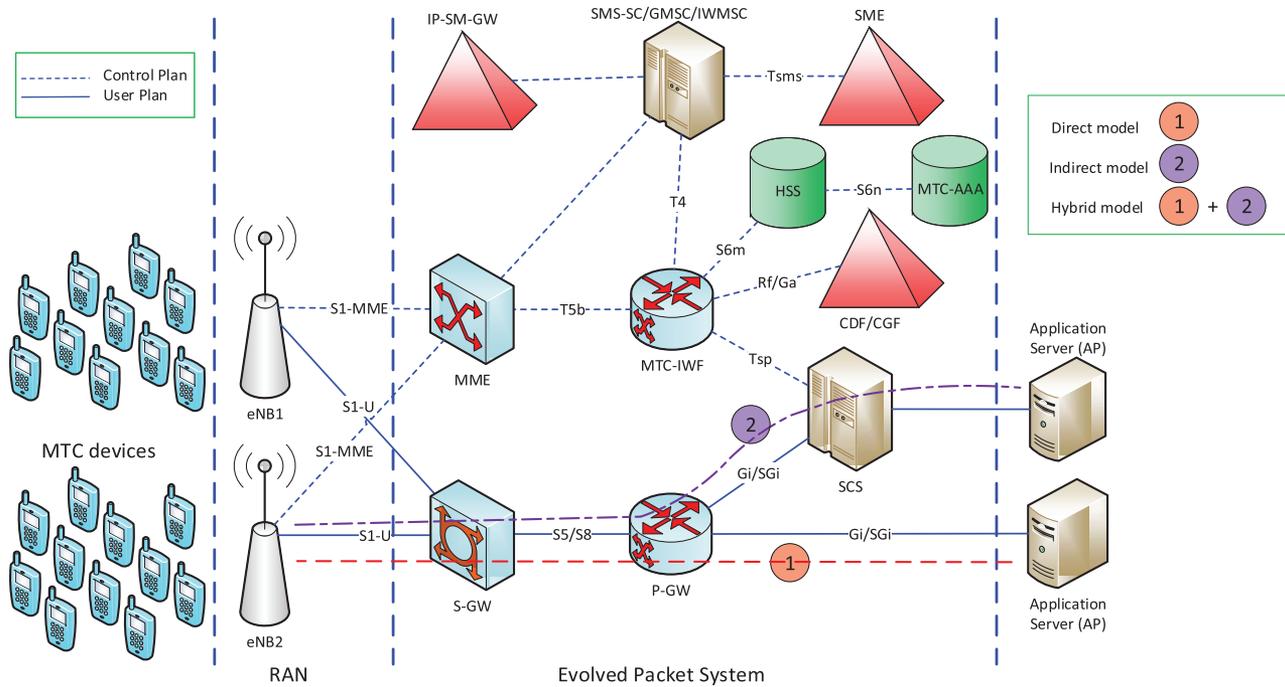


Fig. 1. 3GPP Architecture for Machine-Type Communication [20]

uses the power level that ensures that the signal is well detected by the eNB, without wasting any additional power.

2) *Random Access Response Reception (Msg2)*: Once the random access preamble is transmitted, the terminal monitors the Physical Downlink Control CHannel (PDCCH) to receive the Random Access Response (RAR) message during the RAR window. This message is identified by the Random Access-Radio Network Temporary Identifier (RA-RNTI) associated with the PRACH in which the RA preamble is transmitted. The RAR message consists of the Timing Advance (TA) command, which is used to adjust the uplink synchronization, and the Temporary Cell-Radio Network Temporary Identifier (TC-RNTI). The TC-RNTI is the temporary identity of the terminal in the cell and it is promoted to C-RNTI if the terminal has not yet a one. The RAR message also assigns to the terminal uplink resources to be used in the next step. For Non-contention based RACH procedure, the terminal supposes that the RACH procedure has been successfully finished, while the terminal with contention-based continues to the third step. It is worth noting that the terminals that did not receive a response during the RAR window will do backoff. When the backoff timer expires, they will adjust the power transmission, by the open loop power control, and then retransmit the preamble.

3) *RRC Connection Request (Msg3)*: After the successful reception of Msg2, the terminal adjusts the uplink synchronization and sends the Msg3 containing its ID and the RRC connection request using the UpLink-Shared CHannel (UL-SCH) obtained in the step 2.

4) *RRC Connection Setup (Msg4)*: This step is a response to the precedent one, informing the terminal that RRC connection has been setup. Moreover, this step helps in solving access problems when more than one terminal use the same resources (the same preamble and the same PRACH) while successfully

receiving the second message (*Msg2*). Indeed, the terminals, in this case, share the same temporary identifier (TC-RNTI). Each terminal receiving the downlink message compares the identity in the message with the one transmitted in the third step. Only the terminal observing a match between the two identities will declare that the random access procedure has been successfully finished. After adjusting the power transmission, the other terminals restart the RACH procedure.

C. Power Consumption

As stated earlier, power consumption is very critical for efficient deployment of MTC, especially in case of Massive MTC. The RACH procedure represents one of the most energy consuming procedures in the MTC device lifecycle. Formally speaking, the preamble transmission power can be expressed as follows [30]:

$$P_{PRACH} = \min\{P_{CMAX}, P_{RTP} + PL\} \quad (1)$$

where, P_{CMAX} is the maximum UE transmit power as specified in [31], PL is the Path Loss, explained below. It is worth noting that the maximum value of P_{CMAX} is 23 dBm, as specified by 3GPP. P_{RTP} is the Preamble Received Target Power, which is the perceived power level of the PRACH preamble when reaching the eNB. This power is given by the following equation [32]:

$$P_{RTP} = P_{IRTP} + \Delta_{prmb} + (n_{tr} - 1) * PRS \quad (2)$$

where P_{IRTP} is the Power Initial Received Target Power, representing the initial values by which the PRACH preamble is transmitted for the first time, and it takes the values between (-120 dBm) and (-90 dBm) with a step (2), i.e.

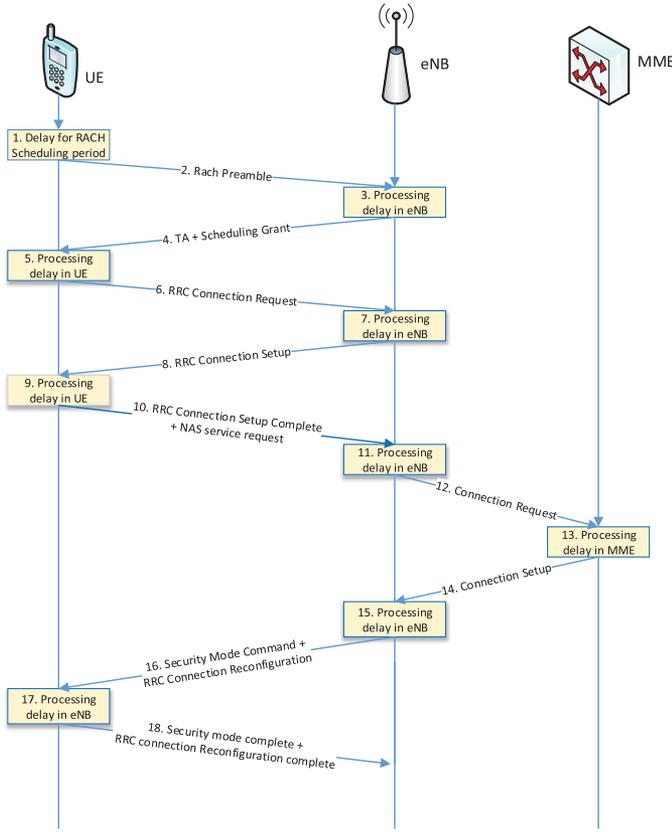


Fig. 2. Control-Plane activation procedure [22]

$PIRTP = \{-120, -118, \dots, -90\}$ dBm. Δ_{prmb} is the preamble format based offset, and its value depends on the preamble format, where $\Delta_{prmb} = 0$ dB for the preamble format 0. n_{tr} is the current number of preamble transmissions. PRS is the Power Ramping Step, which is the power ramping factor, and it can take the following values $\{0, 2, 4, 6\}$ dB [33]. PRS represents the open loop power control during the RACH procedure, wherein the UE increases its transmit power by PRS dB in the next time when the preamble transmission fails. Regarding the pathloss PL , it is the downlink pathloss calculated in the UE in a dB unit. Pathloss can be defined as the signal attenuation between the transmitter and the receiver as function of the propagation distance and other parameters, such as the environment and the frequency [34], [35]. As there is pathloss, the UE should compensate this attenuation so that the signal would reach the receiver with the desired power level.

IV. PROPOSED SOLUTION: FURTHER IMPROVEMENT - TRAFFIC SCATTERING FOR GROUP PAGING (FI-TSFGP)

A. System Model

In the envisioned model, we assume that a group of M MTC devices is distributed over N cells in the network. These devices are uniformly distributed over the cells, and therefore each cell hosts (M/N) MTC devices. Regarding the required resources, the base station, i.e. eNB, reserves R Random Access (RA) channels for the contention-based RACH procedure. Here, the total available resources are defined in terms of Random Access Opportunities (RAOs), which are equal to the number

TABLE I
LIST OF ABBREVIATIONS

Abbreviations	Definition
3GPP	3 rd Generation Partnership Project
5G	5 th Generation
ACB	Access Class Barring
C-RNTI	Cell-Radio Network Temporary Identifier
CDF	Cumulative Distribution Function
CDR	Controlled Distribution of Resources
CN	Core Network
CGP	Consecutive Group Paging
eNB	evolved Node-B
FI-TSFGP	Further Improvement-TSFGP
GID	Group ID
GP	Group Paging
H2H	Human-to-Human
IoT	Internet-of-Things
IR	Initial Ranging
ITS	Intelligent Transport System
LTE-A	Long Term Evolution-Advanced
M2M	Machine-to-Machine
MTC	Machine-Type-Communications
MTC-IWF	MTC-InterWorking Function
PBO	Pre-BackOff
PDCCH	Physical Downlink Control Channel
PLMN	Public Land Mobile Network
PRS	Power Ramping Step
PS	Public Safety
QoS	Quality-of-Services
RA	Random Access
(P)RACH	(Physical)Random Access Channel
RAN	Radio Access Network
RAO	Random Access Opportunity
RAR	Random Access Response
RA-RNTI	Random Access- Radio Network Temporary Identifier
RRC	Radio Resource Control
SCS	Services Capability Server
TC-RNTI	Temporary Cell-Radio Network Temporary Identifier
TSFGP	Traffic Scattering For Group Paging
UE	User Equipment

of frequency bands in the RA slot multiplied by the number of RA preambles. For the sake of simplicity, we suppose that there is just one frequency band, and thus RAOs are equal to the number of RA preambles. After receiving the paging message, addressed by GID, the members of the group will start the contention-based RACH procedure with a certain probability, the activation's probability P_{act} , instead of leaving them to start the RACH procedure all at the same time as in the GP method. P_{act} should ensure that the number of arrivals, new and retransmission attempts, does not exceed an optimal value, in order to maximize the success probability. It is worth noting that the maximum success probability can be achieved when the number of arrivals is equal to the number of channels. P_{act} should also ensure that the number of arrivals at each time does not engender success probability that could not be supported by the network. In other words, the objective is to ensure that the number of successful MTC devices at each time is equal, at most, to the number of MTC devices that can be acknowledged during the Random Access Response (RAR) window. Note that the number of responses during the RAR window is equal to:

$$N_{ACK} = N_{RAR} \times W_{RAR} \quad (3)$$

where N_{RAR} is the maximum number of RARs per a response message.

B. Another Vision of Group Paging

Rather than relying on the good GP analysis presented in [7], we introduce an alternative GP analysis that allows us to well understand our proposed method. After receiving the paging message, all the group members, i.e. M MTC device (by assuming that there is only one cell), start the contention-based RACH procedure in the first available RA slot. After transmitting the preambles, there is a part of MTC devices that successfully transmit the preambles, while the preambles of the others will be collided, not collided but not detected by eNB, or not collided, detected by the eNB, but not indicated by the RAR message. The numbers of successful and collided MTC devices after the first preamble transmission are equal to [7]:

$$M_{i,s} = M_{1,s} = \begin{cases} M e^{-\frac{M}{R}} p_1 & \text{if } M e^{-\frac{M}{R}} p_1 \leq N_{ACK} \\ N_{ACK} & \text{otherwise} \end{cases} \quad (4)$$

$$M_{i,c} = M_{1,c} = M - M_{1,s} \quad (5)$$

where, (i) is the order of the RA slot within the GP interval and p_1 is the detection probability for the first preamble transmission. Generally, for the n^{th} preamble transmission, the probability p_n is equal to $p_n = 1 - e^{-n}$. After finishing the RAR window, all the MTC devices that did not receive a response, i.e. $M_{1,c}$, suppose that a collision has occurred. Therefore, they will do backoff and then restart the RACH procedure by transmitting the preamble once the backoff timer expires. As the backoff time follows a uniform distribution, the collided MTC devices will be uniformly distributed over the next slots during the backoff interval W_{BO} . The number of MTC devices retransmitting their preambles for the next time, in a certain RA slot, is equal to the part of slots (named as α_a , α_{bc} and α_d), from the backoff interval, falling before this RA slot multiplied by the number of collided MTC devices. In the following, we will calculate the position of the RA slots falling within the backoff interval relative to the preamble transmission at the RA slot (i) , and the corresponding proportions, i.e. α_a , α_{bc} and α_d , of the MTC devices whose backoff timers expire and retransmit their preamble at these RA slots. The first RA slot that falls within the backoff window (as illustrated in Fig. 3) will be the one that comes just after the finish of RAR window. It will be at the position:

$$x_a(i) = i + \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil \quad (6)$$

where $x_a(i)$ is the order of the first RA slot within backoff window, relative to the preamble transmission at the RA slot (i) , T_{RAR} is the processing delay at the eNB, and T_{RA_REP} is the interval between two consecutive RA slots. The proportion of the MTC devices whose backoff timers reach zero and hence retransmit their preambles at the RA slot (a) is equal to the time of the slot (a) , in a sub-frame unit, minus the duration before the start of the Backoff window (normalized by W_{BO}):

$$\alpha_a = \frac{[1 + (x_a(i) - 1)T_{RA_REP}]}{W_{BO}} - \frac{[1 + (i - 1)T_{RA_REP} + T_{RAR} + W_{RAR}]}{W_{BO}}$$

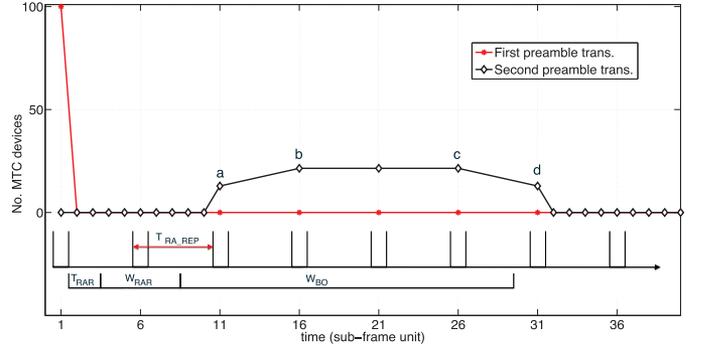


Fig. 3. Number of MTC devices at each RA slot for the first and second preamble transmission for $R = 54$, and $M/N = 100$.

or

$$\alpha_a = \frac{\left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil T_{RA_REP} - (T_{RAR} + W_{RAR})}{W_{BO}} \quad (7)$$

Regarding the RA slots from (b) to (c) , they will be just after the RA slot (a) , i.e.:

$$x_{bc}(i) = x_a(i) + k = i + \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil + k \quad (8)$$

where, $k = 1, 2, \dots, K_{max}$. K_{max} represents the number of RA slots from the backoff window that fall between the slots (b) and (c) . It is equal to (see the appendix B for the proof) $K_{max} = \lfloor (W_{BO} - \alpha_a W_{BO}) / T_{RA_REP} \rfloor$. However, the proportion of MTC devices that retransmit their preambles at these RA slots is equal to:

$$\alpha_{bc} = \frac{T_{RA_REP}}{W_{BO}} \quad (9)$$

The rest of collided MTC devices will transmit their preambles at the last RA slot within the backoff window, i.e. the RA slot (d) . This slot will be just after the last one of the RA slots (bc) , i.e.:

$$\begin{aligned} x_d(i) &= i + \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil + K_{max} + 1 \\ &= i + \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil + \left\lfloor \frac{W_{BO} - \alpha_a W_{BO}}{T_{RA_REP}} \right\rfloor + 1 \\ &= i + \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil + \left\lfloor \frac{W_{BO}}{T_{RA_REP}} - \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil + \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rfloor + 1 \end{aligned}$$

or

$$x_d(i) = i + \left\lfloor \frac{T_{RAR} + W_{RAR} + W_{BO}}{T_{RA_REP}} \right\rfloor + 1 \quad (10)$$

and the proportion of MTC devices in this case is equal to:

$$\begin{aligned} \alpha_d &= 1 - \alpha_a - \alpha_{bc} K_{max} \\ &= 1 - \frac{Q T_{RA_REP} - (T_{RAR} + W_{RAR})}{W_{BO}} - \frac{T_{RA_REP}}{W_{BO}} \left[\frac{W_{BO}}{T_{RA_REP}} - Q + \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right] \end{aligned}$$

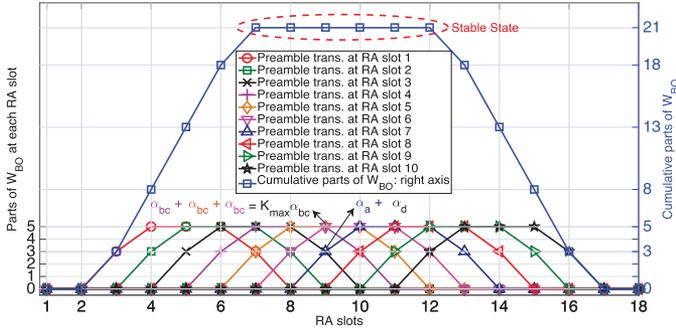


Fig. 4. Cumulative parts of W_{BO} for each RA slot for MTC devices transmitting their preambles for the second time, where $W_{BO} = 21$ and $T_{RA_REP} = 5$

where $Q = \left\lceil \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rceil$. Therefore, α_d is equal to:

$$\alpha_d = \frac{T_{RAR} + W_{RAR} + W_{BO}}{W_{BO}} - \frac{T_{RA_REP}}{W_{BO}} \left\lceil \frac{T_{RAR} + W_{RAR} + W_{BO}}{T_{RA_REP}} \right\rceil \quad (11)$$

It is worth noting that $\alpha_a + K_{max}\alpha_{bc} + \alpha_d = 1$. Accordingly, the numbers of MTC devices retransmitting their preambles for the second time are equal to:

$$M_{retrans} = \begin{cases} M_{1,c} \times \alpha_a & ; \text{for RA slot } a \\ M_{1,c} \times \alpha_{bc} & ; \text{for RA slots } bc \\ M_{1,c} \times \alpha_d & ; \text{for RA slot } d \end{cases} \quad (12)$$

By assuming that each RA slot experiences the same number of new arrivals, the number of successful and collided MTC devices will be the same as given by the equations (4), (5), and (12), and they will generate the same graphic as illustrated in Fig. 3. Therefore, the number of collided MTC devices at each RA slot will be the sum of the contribution of each RA slot, as illustrated in Fig. 4. From this figure, we clearly see that when the number of new arrivals at each RA slot is the same, we come up to a situation where the number of MTC devices retransmitting their preambles is constant. This implies that the number of successful and collided MTC devices at each RA slot will be constant too.

C. Analytical Model

The idea behind our proposition FI-TSFGP is to scatter the MTC devices of a group being paged over the available interval rather than leaving them to start the contention-based RACH procedure all at once. Generally speaking, the number of MTC devices at the RA slot (i) can be written by the following equation:

$$M_i = \sum_{n=1}^{N_{PTmax}} M_i[n] \quad (13)$$

where N_{PTmax} is the maximum number of preamble transmissions, and $M_i[n]$ is the number of MTC devices transmitting their preamble for the n^{th} time in the RA slot (i). The number of successful MTC devices at the RA slot (i) is equal to [7], [8]:

$$M_{i,s}[n] = \begin{cases} M_i[n] e^{-\frac{M_i}{K}} p_n & ; \text{if } M_{toti,s} \leq N_{ACK} \\ \frac{M_i[n] e^{-\frac{M_i}{K}} p_n}{M_{toti,s}} N_{ACK} & ; \text{otherwise} \end{cases} \quad (14)$$

where $M_{toti,s} = \sum_{n=1}^{N_{PTmax}} M_i[n] e^{-\frac{M_i}{K}} p_n$. However, the network can not send back responses to more than N_{ACK} MTC devices even if the number of successful MTC devices is more than N_{ACK} . Hereafter, we will focus on the case where the number of successful MTC devices is less than or equal to N_{ACK} , i.e. $\sum_{n=1}^{N_{PTmax}} M_i[n] e^{-\frac{M_i}{K}} p_n \leq N_{ACK}$. Accordingly, the number of successful MTC devices at the RA slot (i) could be written as:

$$M_{i,s}[n] = M_i[n] e^{-\frac{M_i}{K}} p_n \quad (15)$$

Let M_{arv} denote the number of new arrivals at each RA slot, which represents the value $M_i[1]$, and therefore the number of successful and collided MTC devices will be:

$$M_{i,s}[1] = M_i[1] e^{-\frac{M_i}{K}} p_1 = M_{arv} e^{-\frac{M_i}{K}} p_1 \quad (16)$$

$$M_{i,c}[1] = M_{arv} - M_{i,s}[1] = M_{arv} (1 - e^{-\frac{M_i}{K}} p_1) \quad (17)$$

From Fig. 4, we clearly see that when the total number of MTC devices, and consequently the number of successful MTC devices, is stable (i.e., merely constant), the cumulative parts of W_{BO} is equal to W_{BO} . Therefore, the collided MTC devices, engendered from the precedent RA slots, whose back-off timers expire and retransmit the preamble for the $(n+1)^{th}$ time at the current RA slot, i.e. $M_i[n+1]$, will be equal to the number of collided MTC devices at the current RA slot transmitting their preambles for the n^{th} time, i.e. $M_{i,c}[n]$. This means that $M_{i,c}[n] = M_i[n+1]$. For example, the number of MTC devices transmitting their preamble for the second time is equal to:

$$M_i[2] = \sum_{h=i-H_1}^{i-H_2} \alpha_h M_{h,c}[1] \quad (18)$$

where $H_2 = \lfloor (T_{RAR} + W_{RAR} + W_{BO}) / T_{RA_REP} \rfloor + 1$ and $H_1 = \lceil (T_{RAR} + W_{RAR}) / T_{RA_REP} \rceil$, deduced directly from $x_d(i)$ and $x_a(i)$, respectively. α_h can be one of the following values: α_a , α_{bc} , and α_d . As the system is in the stable state, both the number of collided MTC devices transmitting their preamble for the first time ($M_{h,c}[1]$) and the total number of MTC devices (M_h) are constant. Note that $M_{h,c}[1]$, which is equal to $M_{i,c}[1]$, is given by the equation (17). Then, the equation (18) becomes:

$$M_i[2] = M_{i,c}[1] \times \sum_{h=i-H_1}^{i-H_2} \alpha_h \quad (19)$$

As the cumulative parts of W_{BO} is equal to W_{BO} , we deduce from Fig. 3 and 4 that:

$$\sum_{h=i-H_1}^{i-H_2} \alpha_h = \alpha_a + K_{max}\alpha_{bc} + \alpha_d = 1 \quad (20)$$

and thus $M_i[2] = M_{i,c}[1]$. The numbers of collided and successful MTC devices transmitting their preamble for the second time are equal to:

$$M_i[2] = M_{i,c}[1] = M_{arv}(1 - e^{-\frac{M_i}{R}} p_1)$$

$$M_{i,s}[2] = M_i[2]e^{-\frac{M_i}{R}} p_2 = M_{arv}(1 - e^{-\frac{M_i}{R}} p_1)e^{-\frac{M_i}{R}} p_2$$

$$\begin{aligned} M_{i,c}[2] &= M_i[2] - M_{i,s}[2] \\ &= M_{arv}(1 - e^{-\frac{M_i}{R}} p_1)(1 - e^{-\frac{M_i}{R}} p_2) \\ &= M_{arv} \prod_{k=1}^2 (1 - e^{-\frac{M_i}{R}} p_k) \end{aligned}$$

By induction, we find that:

$$\begin{aligned} M_i[n] &= M_{i,c}[n-1] \\ M_{i,s}[n] &= M_{arv} \prod_{k=1}^{n-1} (1 - e^{-\frac{M_i}{R}} p_k) e^{-\frac{M_i}{R}} p_n \end{aligned} \quad (21)$$

$$M_i[n+1] = M_{i,c}[n] = M_{arv} \prod_{k=1}^n (1 - e^{-\frac{M_i}{R}} p_k)$$

or

$$M_i[n] = M_{i,c}[n-1] = M_{arv} \prod_{k=1}^{n-1} (1 - e^{-\frac{M_i}{R}} p_k) \quad (22)$$

Therefore, the total number of MTC devices at each RA slot, in the stable state, is equal to:

$$M_i = \sum_{n=1}^{N_{PTmax}} M_i[n] = M_{arv} \sum_{n=1}^{N_{PTmax}} \prod_{k=1}^{n-1} (1 - e^{-\frac{M_i}{R}} p_k) \quad (23)$$

The equation (23) can be written by the following form (see the appendix A for the demonstration):

$$M_i = M_{arv} \sum_{m=0}^{N_{PTmax}-1} \alpha_m e^{-\frac{m M_i}{R}} \quad (24)$$

where α_m is:

$$\alpha_m = \sum_{t=1}^{N_{PTmax}-m} (-1)^m \underbrace{\sum_{k_1=1}^t \dots \sum_{k_m=k_{m-1}+1}^{t+m-1} p_{k_1} \dots p_{k_m}}_{m \text{ times}} \quad (25)$$

However, the exponential function can be approximated by the following equation [36]:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \dots \quad (26)$$

Applying this approximation to the equation (24), we find that:

$$\frac{M_i}{M_{arv}} = \sum_{m=0}^{N_{PTmax}-1} \alpha_m - \sum_{m=0}^{N_{PTmax}-1} m \alpha_m \frac{M_i}{R} +$$

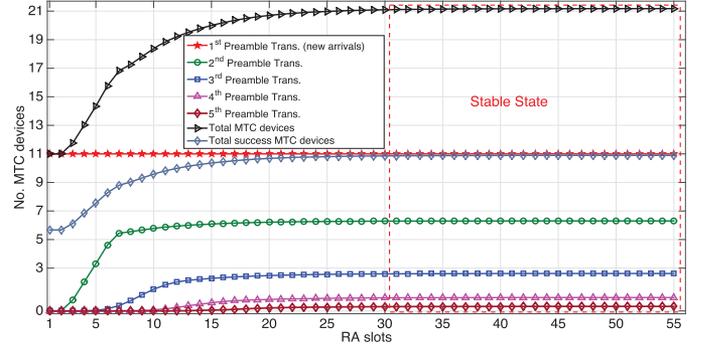


Fig. 5. Number of MTC devices for each preamble transmission as well as the number of total and successful MTC devices in each RA slot; $R = 54$, $N_{PTmax} = 5$

$$\sum_{m=0}^{N_{PTmax}-1} m^2 \alpha_m \frac{M_i^2}{2R^2} \quad (27)$$

or,

$$\begin{aligned} &\left(\sum_{m=0}^{N_{PTmax}-1} m^2 \alpha_m \right) M_i^2 - 2 \left(\frac{R^2}{M_{arv}} + R \sum_{m=0}^{N_{PTmax}-1} m \alpha_m \right) \\ &\times M_i + 2R^2 \sum_{m=0}^{N_{PTmax}-1} \alpha_m = 0 \end{aligned} \quad (28)$$

This equation is a second order one for M_i , which can be solved easily. After obtaining the total number of MTC devices in the stable state, M_i , we calculate the number of successful MTC devices by the following equation:

$$M_{i,s} = \sum_{n=1}^{N_{PTmax}} M_{i,s}[n] \quad (29)$$

where $M_{i,s}[n]$ is given by the equation (21). Fig. 5 shows the number of MTC devices transmitting their preambles for the i^{th} time, and also the total number of arrivals and the number of successful MTC devices ($N_{PTmax} = 5$). It is worth noting that the calculated value by the equation (28) is for the case when the number of arrivals is stable.

Fig. 6-8 and 7-9 illustrate the true and the approximate values of the total number of MTC (equation 28) and the number of successful MTC (equation 29), respectively. These figures include the results for TSFGP as well FI-TSFGP for the sake of comparison. Moreover, different values of R , M_{arv} , and N_{PTmax} were considered. From these figures, we clearly see that TSFGP method generally gives a good estimation of the total number and also the number of successful MTC devices. However, TSFGP fails to estimate the intended values for certain configurations, e.g. $R = 42$, $M_{arv} = 15$ and $N_{PTmax} = 10$ (Fig. 6 and 7). To cope with this shortcoming, FI-TSFGP uses an iterative operation as illustrated in **Algorithm 1**, where δ is the tolerated error. Note that we assume that the value calculated

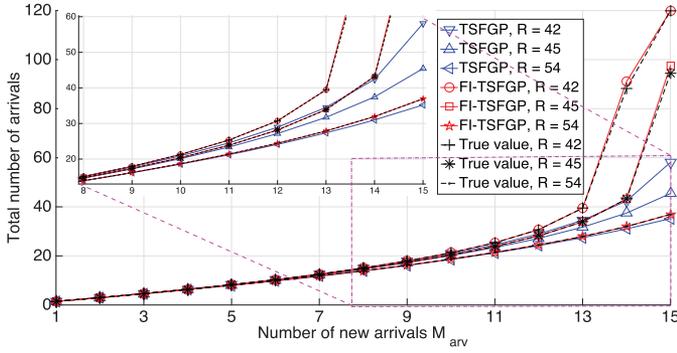


Fig. 6. The total number of arrivals in the stable state as function of the number of new arrivals M_{arv} for different number of preambles; $N_{ACK} = 15$ and $N_{PT_{max}} = 10$

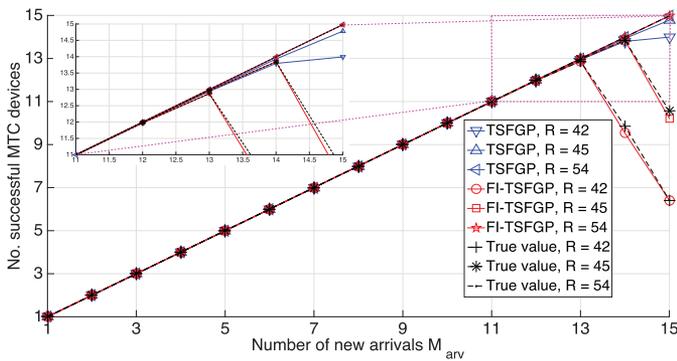


Fig. 7. The number of successful MTC devices in the stable state as function of the number of new arrivals M_{arv} for different number of preambles; $N_{ACK} = 15$ and $N_{PT_{max}} = 10$

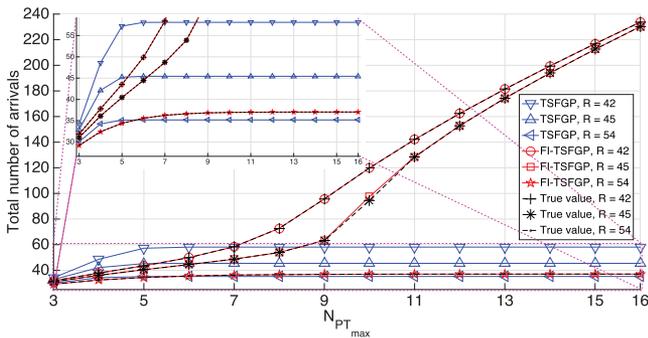


Fig. 8. The total number of arrivals in the stable state as function of the number of preamble transmissions $N_{PT_{max}}$; $M_{arv} = N_{ACK} = 15$

by the equation (28) is the initial guess of the total number of MTC devices in the stable state.

Returning to Fig. 6, 7, 8, and 9, we observe that FI-TSFGP has a great impact when both the number of new arrivals M_{arv} and the number of preamble transmissions $N_{PT_{max}}$ are large. This is attributable to the improvement obtained via the iterative operation. Further, these four figures reveal that FI-TSFGP can be applied for any configuration, while TSFGP is valid for certain configurations. Therefore, FI-TSFGP gives the flexibility to change the network's parameters, e.g. increasing $N_{PT_{max}}$ for increasing the available interval. Regarding the number of successful MTC devices, we remark that, for a fixed value of

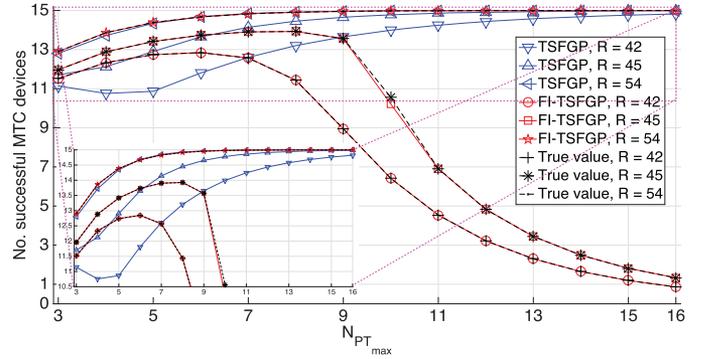


Fig. 9. The number of successful MTC devices in the stable state as function of the number of preamble transmissions $N_{PT_{max}}$; $M_{arv} = N_{ACK} = 15$

Algorithm 1. Iteration operation for further improvement of the approximated value of M_i

$M_{i_{guess}} \leftarrow$ (the solution of equation (28))

$M_{i_{current}} \leftarrow M_{i_{guess}}$

$$M_{i_{new}} \leftarrow M_{arv} \sum_{m=0}^{N_{PT_{max}}-1} \alpha_m e^{-\frac{m M_{i_{current}}}{R}}$$

while $|M_{i_{new}} - M_{i_{current}}| > \delta$ **do**

$M_{i_{current}} \leftarrow M_{i_{new}}$

$$M_{i_{new}} \leftarrow M_{arv} \sum_{m=0}^{N_{PT_{max}}-1} \alpha_m e^{-\frac{m M_{i_{current}}}{R}}$$

end while

$M_{i_{current}} \leftarrow M_{i_{new}}$

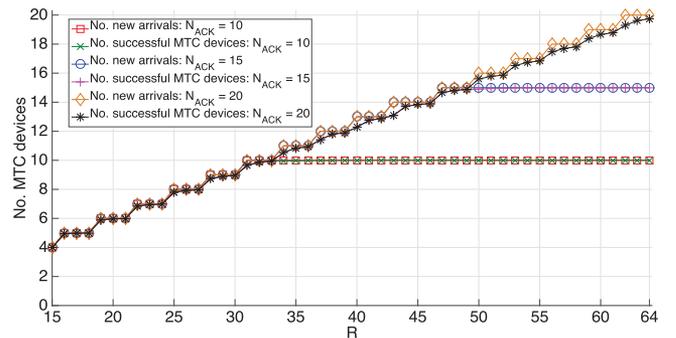


Fig. 10. Number of new arrivals that maximizes the number of successful MTC devices and the corresponding number of successful MTC devices as function of the number of preambles for different values of N_{ACK} ; $N_{PT_{max}} = 10$

$N_{PT_{max}}$, the relationship between the number of new arrivals (M_{arv}) and the number of successful MTC devices is roughly linear as long as M_{arv} is smaller than a certain value, which is equal to ($M_{arv} = 13$) when $R = 42$. Thus, the best number of new arrivals for a certain configuration will be the value that maximizes the number of successful MTC devices as illustrated

in Fig. 10. This figure is highly important since it illustrates the optimal number of new arrivals M_{arv} for a given number of preambles and certain values of N_{ACK} . From this figure, we see that the number of new arrivals (and consequently the number of successful MTC devices) grows as the number of available preambles increases. Moreover, this relationship could be approximated to a linear one. However, when the number of available preambles exceeds certain value ($R = 50$ when $N_{ACK} = 15$), the improvement becomes minimal. In this case, it is more appropriate to choose ($R = 50$) for a better utilization of resources. Taking into account these results, it is better to activate, at each RA slot, a number of MTC devices less than or equal to N_{ACK} , instead of leaving all the members of the group to start the RACH procedure all at the same time. If we need to uniformly distribute (M/N) MTC devices over I_{max} RA slots, there will be, on average, $(M/N)/I_{max}$ MTC device at each RA slot, where I_{max} is given by the following equation [7]:

$$I_{max} = 1 + (N_{PT_{max}} - 1) \left[\frac{T_{RAR} + W_{RAR} + W_{BO}}{T_{RA_REP}} \right]$$

In order to make sure that there will be, on average, M_{arv} MTC devices at each RA slot, we then distribute the devices over a virtual interval containing $I_{V_{max}}$ RA slots, where;

$$I_{V_{max}} = \left\lceil \frac{(M/N)}{M_{arv}} \right\rceil \quad (30)$$

Now, each MTC device randomly generates an integer value between $[1, I_{V_{max}}]$. This value represents the RA slot in which the MTC would start the contention-based RACH procedure. Note that the generated values follow the uniform distribution. If the generated value falls within the interval $[1, I_{max}]$, then this device will start the RACH procedure in this RA slot, otherwise, it goes out and returns to the inactive state. The objective behind this technique is to directly determine whether a MTC will proceed the RACH procedure or not. Thus, we avoid leaving the MTC devices to attempt the transmission at each RA slot, e.g. like the p -persistent mechanism [37]. As the MTC devices are uniformly distributed over the available RA slots, increasing the number of RA slots, i.e. increasing I_{max} , will further improve the performance of FI-TSFGP, where the optimal performance would be achieved when $I_{max} = I_{V_{max}}$. It is worth noting that GP performance can not be improved by increasing the number of RA slots (I_{max}), as all the MTC devices start the RACH procedure at the first available RA slot, by supposing that the number of preamble transmissions ($N_{PT_{max}}$) is fix.

Assuming that the M MTC devices are uniformly distributed over N cells, the FI-TSFGP mechanism could be deployed in a real environment as follows:

1. The network (i.e., eNB) sends the paging message to the intended MTC devices, containing the number of MTC devices to be paged (M/N) and indicating the maximum number of new arrivals, M_{arv} , that the network can support at each RA slot.
2. When receiving the paging message, the MTC device can calculate the virtual interval $I_{V_{max}}$, via the equation (30), using the values sent in the received message.

TABLE II
BASIC SIMULATION PARAMETERS

Notations	Definition	Values
M/N	Average number of devices in each cell	10–5000
R	Total number of preambles in a random access slot	54
BI	Backoff indicator in a sub-frame unit	20
$N_{PT_{max}}$	Maximum number of preamble transmissions	16
N_{RAR}	Maximum number of RARs that can be carried in one response message	3
T_{RAR}	Processing delay required by the eNB in order to detect the transmitted preamble in a sub-frame unit	2
W_{RAR}	Size of random access response window in a sub-frame unit	5
N_{ACK}	Maximum number of MTC devices that can be acknowledged within the RAR window	$N_{ACK} = N_{RAR} \times W_{RAR}$
$PRACH_{conf\{q\}_ind\{x\}}$	PRACH configuration index	$PRACH_{conf\{q\}_ind\{x\}} = 6$
T_{RA_REP}	The interval between two consecutive RA slots	5
W_{BO}	Backoff window size	$BI + 1$
p_n	Preamble detection probability for the n -th preamble transmission	$p_n = 1 - e^{-n}$
T_{CRT}	Contention Resolution timer	48
P_{HARQ_RET}	HARQ retransmission probability for Msg3 and Msg4 (non-adaptive HARQ)	10%
N_{HARQ}	Maximum number of HARQ TX for Msg3 and Msg4 (non-adaptive HARQ)	5
Parameters relative to CGP		
C_{max}	Number of paging cycles	7
$N_{PT_{max}}$	Maximum number of preamble transmissions	3
Parameters relative to PBO		
W_{PBO}	Pre-backoff window in a sub-frame unit	240
$N_{PT_{max}}$	Maximum number of preamble transmissions	8
Parameters relative to Power consumption		
P_{cMAX}	Maximum transmit power	0 dBm
P_{IRTP}	Preamble initial received target power	-120 dBm
Δ_{prmb1}	Preamble format based offset	0 dB
PRS	Power Ramping Step	0 dB
PL	Pathloss	98 dBm
P_0	Power consumption in the inactive state	0 mW
P_1	Power consumption when the UE is waiting for RA slot, and also when it is in backoff	-37 dBm
P_2	Power consumption when the UE receives (or expected to receive) the messages Msg2 and Msg4	-22 dBm
P_3	Power consumption when the UE transmits a signal, such as preamble transmission	$P_3 = P_{PRACH}$

3. Regarding the interval I_{max} , it can be either calculated by the MTC device using the parameters broadcasted by the network or explicitly sent in the paging message.
4. After obtaining the values I_{max} and $I_{V_{max}}$, the MTC device generates an integer value $k \in [1, I_{V_{max}}]$.
5. If $k \leq I_{max}$, the MTC device starts the contention-based RACH procedure at the k^{th} RA slot. Otherwise, it disconnects and changes to the inactive state.

Looking at the precedent steps of FI-TSFGP, it becomes apparent that FI-TSFGP is simple to be applied in practice. Compared to the original Group Paging (GP) method, FI-TSFGP incurs at MTC devices only minimal additional calculations as per steps 2), 4), and 5).

V. PERFORMANCE EVALUATION

In order to evaluate the performance of FI-TSFGP, we built a C++-based discrete events simulator. In the simulation, a group of MTC devices ranging from (10) to (5000) has been considered. Regarding the parameters of RACH procedure, we used those specified in Table 6.2.2.1 in [15]. Furthermore, the control plane latency analysis will be taken as specified in Table B.1.1.1 – 1 of [22]. For the sake of simplicity, the pathloss remains constant and is the same for all the MTC devices. Table II summarizes the parameters used in our study. Regarding the parameters of power consumption (Table II), P_2 is taken to be P_3 for the first time of preamble transmission and P_1 is about 30 times less than P_2 . The power ramping factor is set to be zero (i.e., it will be nullified). As the average waiting time for the first available RA slot is equal to

$T_{RA_REP}/2 = 2.5$ ms, the average power consumption during this period is equal to $2.5 \times 10^{0.1 \times P_1}$ mW, which is added for all the MTC devices at the start of the simulation. In order to show how our proposition behaves, FI-TSFGP will be compared with the Group Paging (GP), Consecutive Group Paging (CGP) [10], and Pre-BackOff (PBO) methods [11]. The main idea of CGP is to repeat the paging interval many times so that the MTC devices having not succeeded in the first paging interval will try to access in the next paging interval and so on. Note that the number of paging cycles, i.e. C_{max} , and $N_{PT_{max}}$ are chosen to be equal to (7) and (3), respectively, as these values maximize the performance of CGP [10]. Regarding the PBO method [11], all the members of the intended group will do backoff before the first preamble transmission. As the authors did not give the optimal value of pre-backoff for a group size more than 1000 MTC devices, we choose an integer value (using the equation 1 of [11]) higher than 200. Indeed, the optimal value of PBO for a group size of 1000 MTC devices is more than 200. Then, the chosen value is $W_{PBO} = 240$ and the corresponding maximum number of preamble transmissions is $N_{PT_{max}} = 8$. We recall that FI-TSFGP tries to activate, at each RA slot during the available interval, the number of MTC devices that maximizes the performance. However, PBO tries to spread the MTC devices during a certain interval regardless the size of the intended group.

A. Performance Metrics

The metrics considered to evaluate the performance of the four above-mentioned schemes are: success, collision, and drop (only for FI-TSFGP) probabilities, average access delay, average number of preamble transmissions, resource utilization, Cumulative Distribution Function (CDF) of both preamble transmission and access delay, and power consumption. The success probability is defined as the number of MTC devices successfully finished their RACH procedure within the maximum number of preamble transmissions, normalized by the total MTC devices (activated and non-activated, for FI-TSFGP, ones). The collision probability is the ratio between the number of collided RAOs and the total number of available RAOs. Since M_{arv} MTC devices will be activated at each RA slot for FI-TSFGP, there will be a part of MTC devices that will not be activated when $((M/N) > I_{max}M_{arv})$. Thus, the drop's probability is equal to:

$$P_d = \begin{cases} \frac{(M/N) - I_{max}M_{arv}}{(M/N)} & \text{if } (M/N) > I_{max}M_{arv} \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

Regarding the average access delay, it represents the total access delay for all the MTC devices, which successfully finished the RACH procedure, between the first preamble transmission and the completion of the random access procedure (within the maximum number of preamble transmissions) divided by the total number of successful MTC devices [6], [15]. The average number of preamble transmissions is the total number of preamble transmissions of all the MTC devices successfully finished the RACH procedure, divided by the number of successful MTC devices. For CGP scheme, this time will

be the sum of the access delay in the current paging interval plus the time of the precedent paging intervals, and the same thing for the average number of preamble transmissions. Let (r) be the number of preamble transmissions, then the CDF of preamble transmission can be defined as the number of MTC devices successfully finished their RACH procedure by (r) times or less of preamble transmission divided by the total number of successful MTC devices, and it is given by the following equation:

$$CDF_R(r) = \frac{\sum_{i=1}^{I_{max}} \sum_{n=1}^r M_{i,s}[n]}{\sum_{i=1}^{I_{max}} \sum_{n=1}^{N_{PT_{max}}} M_{i,s}[n]} \quad (32)$$

Let (d) be the access delay for the RACH procedure between the first attempt and the completion of the RACH procedure. The CDF of access delay can be defined as the number of MTC devices successfully finished the RACH procedure before the time (d) and the total number of successful MTC devices. It is given by the following equation:

$$CDF_D(d) = \frac{\sum_{t=1}^d M_{s,t}}{\sum_{t=1}^{T_{max}} M_{s,t}} \quad (33)$$

where $M_{s,t}$ is the number of successful MTC devices whose access delay is equal to (t) , and T_{max} is the maximum access delay that is equal to the time of the paging interval in a sub-frame unit, i.e. $T_{max} = 1 + (I_{max} - 1) * T_{RA_REP} + T_{RAR} + W_{RAR}$. The resource utilization (RU) can be defined as the ratio of the total number of successful MTC devices to the total available RAOs, and it can be given by the following equation:

$$RU = \frac{\sum_{i=1}^{I_{max}} \sum_{n=1}^{N_{PT_{max}}} M_{i,s}[n]}{I_{max}R} \quad (34)$$

Regarding the power consumption, four values are considered: the power consumption for successful, collided, dropped (just for FI-TSFGP), and the total number of MTC devices. The Power consumption for successful/collided/dropped MTC devices is the mean power consumption of the MTC devices having successfully accessed the network/collided/dropped, respectively. These parameters will be calculated for GP method, and then generalized for FI-TSFGP. Usually, the power consumption of the successful MTC devices consists of the following parts (we assume that the device needs n preamble transmission before a successful attempt):

1. The power consumption when the device is waiting for the first RA slot; $(T_{RA_REP}/2)P_1$.
2. The power consumption when the device is transmitting the preamble for the first $(n - 1)$ times, and collision occurs. This power is equal to the one consumed in the following steps; transmitting the preamble (P_3), waiting for the RAR window ($T_{RAR}P_1$), during the RAR window ($W_{RAR}P_2$), and during the backoff and waiting for the next RA slot ($(\lceil(1 + T_{RAR} + W_{RAR} + W_{BO}/2)/T_{RA_REP}\rceil T_{RA_REP} - 1 - T_{RAR} - W_{RAR})P_1$). Note that $W_{BO}/2$ is the average time of the backoff as the backoff timer can expire at any time during the backoff window.

3. The power consumed during the n^{th} preamble transmission (the successful transmission); which is equal to $P_3 + T_{RAR}P_1 + (W_{RAR}/2)P_2$. $(W_{RAR}/2)P_2$ is the average power consumption during the RAR window as the MTC device can receive the RAR message at any sub-frame during the RAR window.
4. The power consumed for the messages Msg3 and Msg4 of the RACH procedure (by ignoring the effect of Msg3 and Msg4 retransmission [7]), which is the power consumed during the processing of the message Msg2 ($T_{P_{Msg2}}P_2$), the power consumed during the transmission of Msg3 (P_3), the power consumed when the MTC is waiting for the acknowledgment (ACK) of Msg3 ($T_{HARQ}P_2$), the power consumed when receiving the ACK of Msg3 (P_2), the power consumed when receiving the Msg4 (P_2), the power consumed after receiving Msg4 and before transmitting the ACK of Msg4 ($T_{HARQ}P_2$), and finally the power consumed for transmitting the ACK of Msg4 (P_3).

Accordingly, the power consumption for the MTC devices successfully accessed the network is:

$$W_S = \frac{T_{RA_REP}}{2} P_1 + (n-1)(P_3 + T_{RAR}P_1 + W_{RAR}P_2 + (\lceil \frac{1+T_{RAR}+W_{RAR}+W_{BO}/2}{T_{RA_REP}} \rceil T_{RA_REP} - 1 - T_{RAR} - W_{RAR})P_1) + P_3 + T_{RAR}P_1 + (W_{RAR}/2)P_2 + T_{P_{Msg2}}P_2 + P_3 + T_{HARQ}P_2 + P_2 + T_{HARQ}P_2 + P_3 \quad (35)$$

or

$$W_S = (T_{RA_REP}/2 + (n-1)\lceil \frac{1+T_{RAR}+W_{RAR}+W_{BO}/2}{T_{RA_REP}} \rceil) \times T_{RA_REP} + T_{RAR} - (n-1)W_{RAR} - (n-1)P_1 + (1 + (n-1/2)W_{RAR} + 2T_{HARQ} + T_{P_{Msg2}})P_2 + (n+2)P_3 \quad (36)$$

As regards to the power consumption of the failed MTC devices, it can be deduced directly from the average power consumption of the successful MTC devices, wherein the number of preamble transmissions is the maximum allowed one and there is no transmission of Msg3 and Msg4. Therefore, it is equal to:

$$W_F = \frac{T_{RA_REP}}{2} P_1 + (N_{PT_{max}} - 1)(P_3 + T_{RAR}P_1 + W_{RAR}P_2 + (\lceil \frac{1+T_{RAR}+W_{RAR}+W_{BO}/2}{T_{RA_REP}} \rceil T_{RA_REP} - 1 - T_{RAR} - W_{RAR})P_1) + P_3 + T_{RAR}P_1 + W_{RAR}P_2 \quad (37)$$

The average power consumption of the dropped MTC devices (only for FI-TSFGP) is equal to $W_D = \frac{T_{RA_REP}}{2} P_1 + (I_{max} - 1)T_{RA_REP}P_0$. The Power consumption for the total number of MTC devices is the mean power consumption of all the MTC devices, i.e. successful, failed, and dropped, and it is given by the following equation:

$$W = \frac{M_S W_S + M_F W_F + M_D W_D}{M_S + M_F + M_D} \quad (38)$$

where M_S , M_F , and M_D are the number of successful, failed, and dropped MTC devices. For the power consumption of FI-TSFGP, it is sufficient to add the value $(k-1)T_{RA_REP}P_1$, where $k \in [1, I_{max}]$, as the MTC device is waiting for its RA slot identified by the value k .

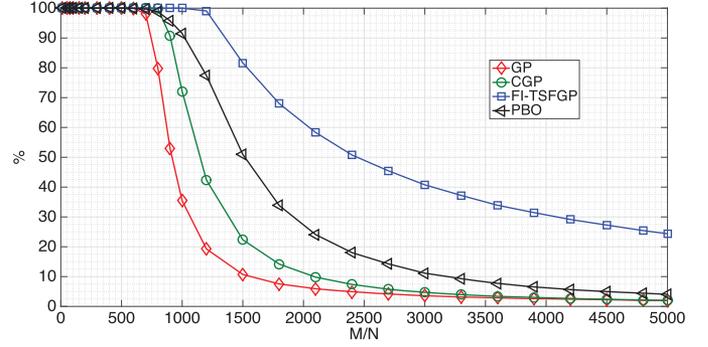


Fig. 11. The success probability for the considered methods

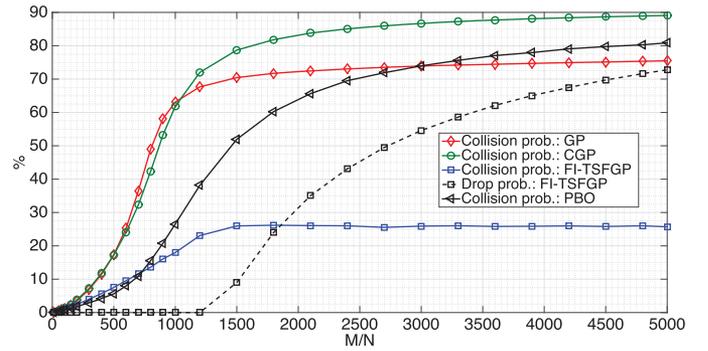


Fig. 12. The collision and drop probabilities for the considered methods

B. Results

Fig. 11 illustrates the success probability of the four considered methods, i.e. GP, CGP, PBO, and FI-TSFGP. We observe that CGP introduces an important improvement, compared to GP, when the number of MTC devices in the group is moderate (nearly until 2500). However, the behavior of CGP becomes similar to that of GP when the number of MTC devices becomes larger than 2500, for the considered parameters. For PBO, we observe that it outperforms both GP and CGP, regardless the size of the group. However, the success probability becomes small when there is a large number of MTC devices. Concerning FI-TSFGP, we clearly see that there is a large improvement, even when the number of MTC devices in each group is large. We note that the success probability for FI-TSFGP is more than 20% when the number of MTC devices is large (e.g., 5000), while it is less than 5% for PBO. Furthermore, the collision probability of FI-TSFGP, as illustrated in Fig. 12, slightly increases as the number of MTC devices increases, and then remains roughly stable below 30%, while this probability is more than 70% for GP and more than 85% for CGP. This means that FI-TSFGP achieves a degradation to about the third. Comparing with FI-TSFGP, we see that the collision probability of PBO is similar for a small size of group, with a little improvement brought by PBO. By increasing the number of MTC devices, the collision probability of PBO keeps increasing, and it becomes even worse than GP for a large number of MTC devices (more than 3000 MTC devices for the considered parameters). We remark that CGP achieves an important improvement regarding the average access delay (Fig. 13) and the average number of preamble

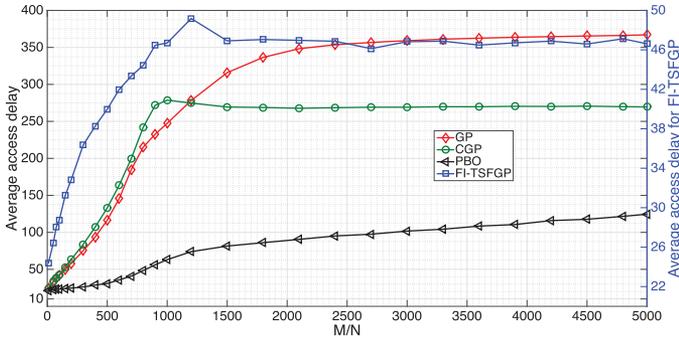


Fig. 13. The average access delay for the considered methods

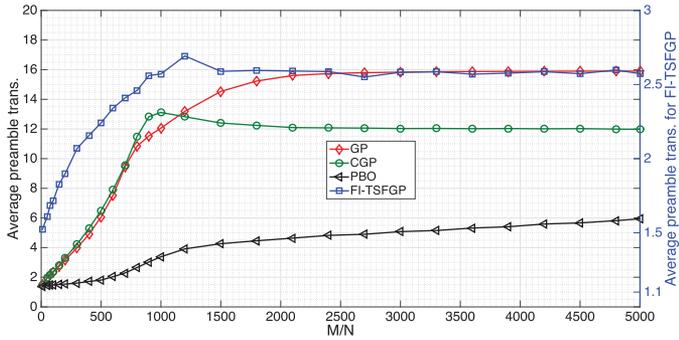


Fig. 14. The average preamble transmission for the considered methods

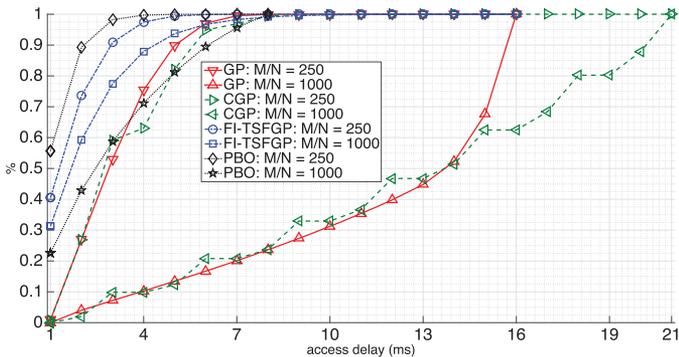


Fig. 15. CDF of Preamble transmissions

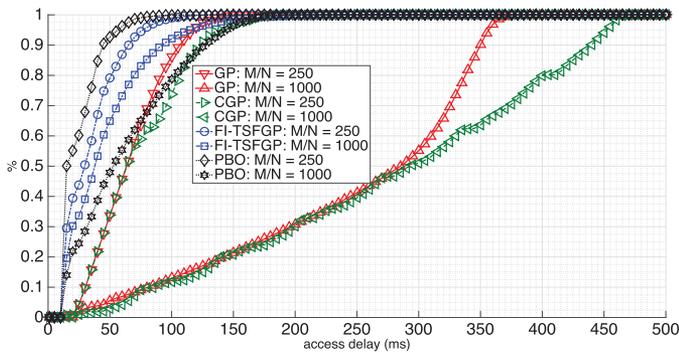


Fig. 16. CDF of access delay

transmissions (Fig. 14) by report to GP. However, there is some price on this improvement, given that a part of MTC devices has access delay and number of preamble transmissions higher than in GP method, (Fig. 15 and 16). For PBO, We remark

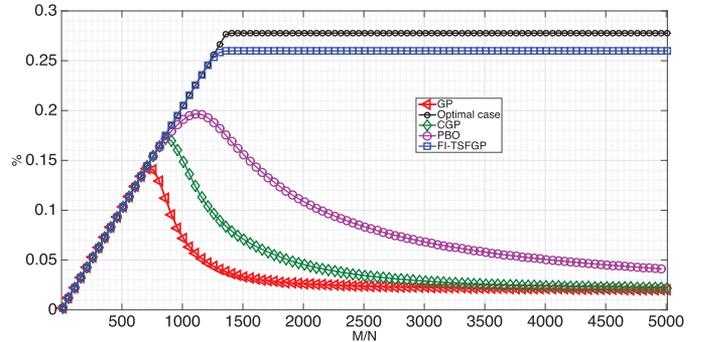


Fig. 17. The resource utilization for the considered methods

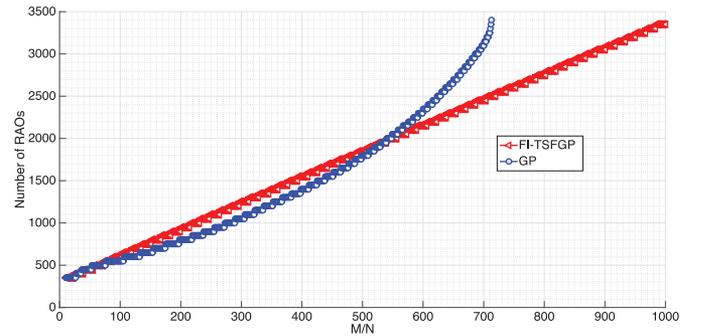


Fig. 18. The minimum resources required in order to achieve 90% of success probability

that it improves these performances by considerably reducing the average access delay and the average number of preamble transmissions. We see, again, that the average access delay and the average number of preamble transmissions of FI-TSFGP is similar to that of PBO for a small group size, with a small improvement brought by PBO. However, FI-TSFGP's performances become better when increasing the number of MTC devices. An important observation can be also seen from Fig. 13 and 14, wherein the average (access delay/preamble transmission) of FI-TSFGP becomes constant after certain size of the group (more than 1500 for the considered parameters), while these values are increasing with the number of MTC devices for PBO. Regarding the CDF of (access delay/preamble transmission), we see that PBO outperforms GP and CGP regardless the number of MTC devices, while it outperforms FI-TSFGP only for the case of small group sizes. However, FI-TSFGP outperforms all the considered methods, including PBO, for larger group sizes, where the achieved gain can reach more than 15% for CDF of preamble transmission and more than %40 for CDF of access delay. It should be noted that the number of preamble transmissions needed to access the network and thus the time required to get access have a close relation with the power consumption. Therefore, FI-TSFGP introduces a large reduction of the power consumption, which is a very important achievement, especially for those with a limited power resources. Furthermore, FI-TSFGP largely reduces the access delay, which is an important issue for the time-critical MTC applications, for example.

Looking at the resource utilization, Fig. 17 shows again that the CGP achieves some improvement when the number of MTC

devices is somewhat moderate, while the behavior becomes nearly the same as of GP when the number of MTC devices is large. As for the precedent performance metrics, PBO method has a better resource utilization, compared to GP and CGP. But, this utilization decreases when exceeding the number of MTC devices after $1000 M/N$. FI-TSFGP achieves a high percent of resource utilization, similar to that of the ideal case. Note that the latter represents the situation where the total number of arrivals engenders a number of successful MTC devices that is equal to N_{ACK} , i.e. the number of MTC devices that the network can acknowledge within the RAR window. Furthermore, we remark that there is a small difference between FI-TSFGP and the ideal case, which is about 2.5% when the number of MTC devices is large, while it is more than 20% for GP, CGP, and PBO. From Fig. 17, we also observe that FI-TSFGP maintains a stable resource utilization regardless the size of the group. This means that FI-TSFGP achieves a constant number of successful MTC devices whatever the group size, while the other methods fail to do that. Another improvement gained by FI-TSFGP is the minimum resources to achieve 90% of success probability. Fig. 18 shows the relationship between the required resources to achieve 90% of success probability and the number of MTC devices. From this figure, we see that the required resources for FI-TSFGP is more than that for GP, when the number of MTC devices is small. To better explain this behavior we return to Fig. 5, where we clearly see that the number of successful MTC devices, at the start of the group paging interval, is not equal to that value in the stable state, when there is a small number of MTC devices, the average number of successful MTC devices at each RA slot will be relatively low (compared to the reserved resources). Generally, the higher is the number of MTC devices, the higher is the average number of successful MTC devices. However, the relation shown in Fig. 18 can be approximated to an exponential one for GP, and a linear one for FI-TSFGP. This advantage is very important, as we can achieve the same percentage of success with much more less of resources, especially with the existence of a very large number of MTC devices.

Figs. 19 and 20 illustrate the power consumption of the successful MTC devices and that of the total number of MTC devices, and the power consumption of the failed MTC devices and the dropped (only for FI-TSFGP) ones, respectively. From Fig. 19, we observe that the power consumption of successful MTC devices for CGP is smaller than that of GP. This is expected as the average number of preamble transmissions of CGP is smaller than that of GP in the presence of a large number of MTC devices. However, GP outperforms CGP when considering the power consumption of failed and total number of MTC devices (Figs. 19 and 20). We argue this by the fact that the collision probability and the total number of preamble transmissions are larger for CGP, where the total number of preamble transmissions is 21 for CGP and 16 for GP, as shown in Fig. 15. For the same reasons, PBO highly outperforms GP and CGP regarding the three considered values of power consumption, and thus highly conserves the energy. We remark that FI-TSFGP highly outperforms both GP and CGP for all the considered values. Compared with PBO, the power consumption of FI-TSFGP is similar for small group sizes (with a small difference), while FI-TSFGP outperforms PBO for all the

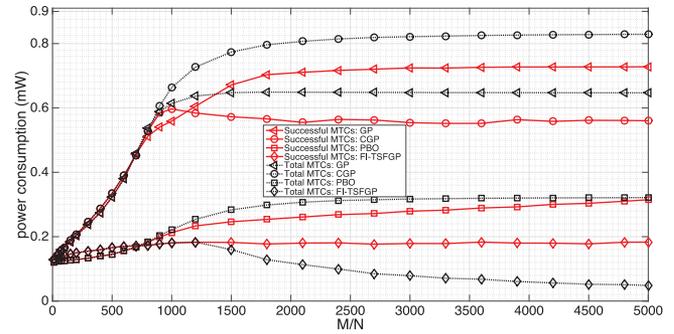


Fig. 19. Power consumption of the successful MTC devices and that of the total MTC devices

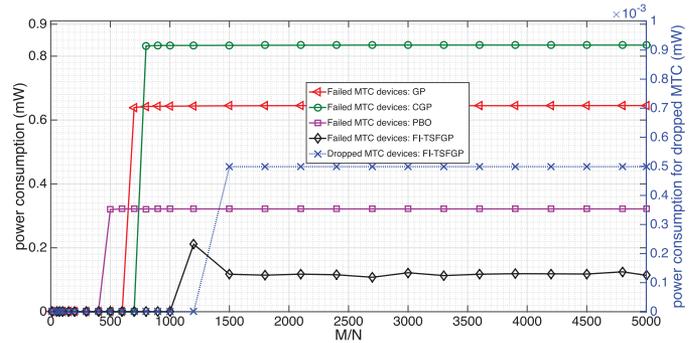


Fig. 20. Power consumption of the failed and dropped MTC devices

considered values when increasing the number of MTC devices. From Fig. 19, we clearly see that the power consumption for successful MTC devices is less than 0.2 mW for FI-TSFGP, while it is about 0.55 mW for CGP, more than 0.70 mW for GP, and about 0.30 mW for PBO for a large number of MTC devices. Another important improvement is the average power consumption of the total MTC devices. From Fig. 19, we clearly observe that the average power consumption for GP, CGP, and PBO increases as the number of MTC devices increases, and then it becomes merely stable (or so slowly increases) when the number of MTC devices becomes large. These values is about 0.65 mW for GP, more than 0.80 mW for CGP, and about 0.3 mW for PBO. However, the average power consumption for FI-TSFGP firstly increases as the number of MTC devices increases, and then it decreases. The decreasing behavior of the average power consumption for FI-TSFGP can be justified by the fact that there is a part of MTC devices that are dropped, i.e. they come back to inactive state. Indeed, when go idle, the dropped devices consume a very small amount of power, compared to the activated ones, and their numbers would be increased when increasing the number of MTC devices. Therefore, the average power consumption of the total number of MTC devices logically decreases as the number of MTC devices increases.

To further show the effectiveness of FI-TSFGP, Fig. 21 and 22 illustrate the CDF of the power consumption for the successful MTC devices and the total number of MTC devices, respectively. From Fig. 21, we observe the superiority of PBO compared to GP and CGP, while it introduces some improvement by report to FI-TSFGP only for small group sizes. For a large number of MTC devices (e.g., 5000), we clearly see that more than 90% of the MTC devices consume only

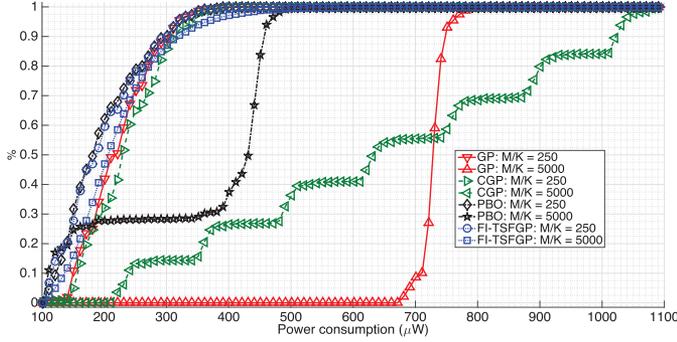


Fig. 21. CDF of power consumption for the successful MTC devices

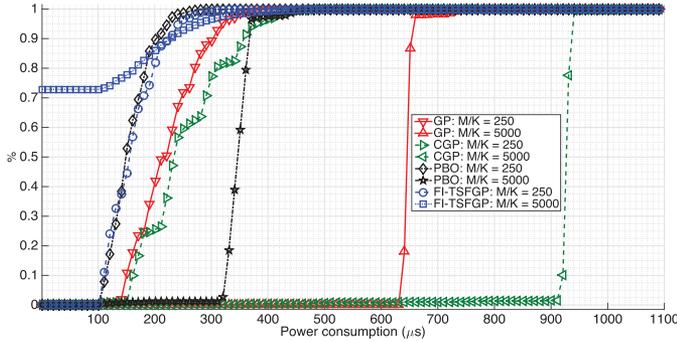


Fig. 22. CDF of power consumption for the total number of MTC devices

0.3 mWatt for FI-TSFGP method, while they consume more than 0.7 mWatt for the GP method, more than 1 mWatt for the CGP method, and more than 0.4 mWatt for PBO method.

Besides the superiority of FI-TSFGP, shown by Fig. 22 for the CDF of power consumption of the total number of MTC devices (compared to GP and CGP and for a large number of MTC devices by report to PBO), we also see a behavior specific to the FI-TSFGP method. This behavior is that the percentage of the total number of MTC devices consuming certain power level augments as the number of MTC devices increases. We justify this benefit of FI-TSFGP by the fact that the network activates certain number of MTC devices, while the others return back to the inactive state in which the MTC devices consume the minimum power level. Therefore, the higher number of MTC devices is, the higher percentage of MTC devices consuming a certain power level is. Taking into account the fact that the number of MTC devices is naturally large, we conclude that our proposed method FI-TSFGP outperforms the other methods for all the considered parameters. Based on the aforementioned results, especially the ones concerned the power consumption, we believe that our method is very attractive for battery-limited MTC devices deployment.

VI. CONCLUSION

In this paper, FI-TSFGP method has been proposed in order to improve the performance of group paging. Our proposition has been evaluated for a relatively large number of MTC devices (5000 MTC devices). FI-TSFGP has outperformed both the Group Paging (GP) and the Consecutive Group Paging (CGP) methods, for all the considered metrics. Compared with PBO, FI-TSFGP method has a similar performance for a low

number of MTC devices, while it outperforms PBO for a large number of MTC devices. Besides the access delay and the average number of preamble transmissions improvements, FI-TSFGP highly reduces the power consumption for both the successful MTC devices and also for the total number of MTC devices, which is one of the key objectives of 5G systems. Moreover, FI-TSFGP maintains a stable resource utilization when existing a large number of MTC devices, meaning that the number of successful MTC devices is maintained regardless the group size. Finally, FI-TSFGP gets the same percentage of success probability for MTC with a much more less of resources, preserving thus the network resources, which can be used by Non-MTC devices, for example.

APPENDIX A

THE PROOF OF THE EQUATION 24

In this section, we try to rewrite the equation (23). First of all, we have

$$W_i = \frac{M_i}{M_1} = \sum_{n=1}^{N_{PTmax}} W_i[n] = \sum_{n=1}^{N_{PTmax}} \left(\prod_{k=1}^{n-1} (1 - e^{-\frac{M_i}{R} p_k}) \right)$$

When varying n from 1 to N_{PTmax} , we have

$$W_i[1] = 1$$

$$W_i[2] = 1 - e^{-\frac{M_i}{R} p_1}$$

$$W_i[3] = (1 - e^{-\frac{M_i}{R} p_1})(1 - e^{-\frac{M_i}{R} p_2}) \\ = 1 - (p_1 + p_2)e^{-\frac{M_i}{R}} + p_1 p_2 e^{-\frac{2M_i}{R}}$$

$$W_i[4] = (1 - e^{-\frac{M_i}{R} p_1})(1 - e^{-\frac{M_i}{R} p_2})(1 - e^{-\frac{M_i}{R} p_3}) \\ = 1 - (p_1 + p_2 + p_3)e^{-\frac{M_i}{R}} + (p_1 p_2 + p_1 p_3 \\ + p_2 p_3)e^{-\frac{2M_i}{R}} - p_1 p_2 p_3 e^{-\frac{3M_i}{R}}$$

$$W_i[5] = (1 - e^{-\frac{M_i}{R} p_1})(1 - e^{-\frac{M_i}{R} p_2})(1 - e^{-\frac{M_i}{R} p_3}) \\ (1 - e^{-\frac{M_i}{R} p_4})$$

$$= 1 - (p_1 + p_2 + p_3 + p_4)e^{-\frac{M_i}{R}} + (p_1 p_2 + p_1 p_3 + \\ p_1 p_4 + p_2 p_3 + p_2 p_4 + p_3 p_4)e^{-\frac{2M_i}{R}} - (p_1 p_2 p_3 + \\ p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4)e^{-\frac{3M_i}{R}} + \\ p_1 p_2 p_3 p_4 e^{-\frac{4M_i}{R}}$$

⋮

Now, if we try to make the sum for the similar terms, we can find that:

$$W_i = W_i[1] + W_i[2] + W_i[3] + W_i[4] + W_i[5] + \dots \\ = \sum_{t=1}^{N_{PTmax}-0} (-1)^0 1 + \sum_{t=1}^{N_{PTmax}-1} (-1)^1 \left(\sum_{k_1=1}^t p_{k_1} \right) \\ \times e^{-\frac{1 \times M}{R}} + \sum_{t=1}^{N_{PTmax}-2} (-1)^2 \left(\sum_{k_1=1}^t \sum_{k_2=k_1+1}^{t+1} p_{k_1} p_{k_2} \right) e^{-\frac{2 \times M}{R}} \\ + \sum_{t=1}^{N_{PTmax}-3} (-1)^3 \left(\sum_{k_1=1}^t \sum_{k_2=k_1+1}^{t+1} \sum_{k_3=k_2+1}^{t+2} p_{k_1} p_{k_2} p_{k_3} \right) \\ e^{-\frac{3 \times M}{R}} + \dots$$

(39)

From the equation (39), we can conclude that:

$$W_i = \sum_{m=0}^{N_{PT_{max}}-1} \sum_{t=1}^{N_{PT_{max}}-m} (-1)^m \times \underbrace{\sum_{k_1=1}^t \sum_{k_2=k_1+1}^{t+1} \dots \sum_{k_m=k_{m-1}+1}^{t+m-1}}_{m \text{ times}} p_{k_1} \dots p_{k_m} e^{-\frac{mM_i}{R}} \quad (40)$$

Let α_m be equal to:

$$\alpha_m = \sum_{t=1}^{N_{PT_{max}}-m} (-1)^m \sum_{k_1=1}^t \dots \sum_{k_m=k_{m-1}+1}^{t+m-1} p_{k_1} \dots p_{k_m} \quad (41)$$

Therefore, we have:

$$W_i = \frac{M_i}{M_1} = \sum_{m=0}^{N_{PT_{max}}-1} \alpha_m e^{-\frac{mM_i}{R}} \quad (42)$$

which is equal to the equation (24).

APPENDIX B

THE PROOF OF K_{max}

In this appendix, we will try to prove whether the value K_{max} is true. It is given in the text by the following equation

$$K_{max} = \left\lfloor \frac{W_{BO} - \alpha_a W_{BO}}{T_{RA_REP}} \right\rfloor = \left\lfloor \frac{T_{RAR} + W_{RAR} + W_{BO}}{T_{RA_REP}} \right\rfloor - \left\lfloor \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rfloor$$

Let $T_{RAR} + W_{RAR} + W_{BO} = \Psi$ and $T_{RAR} + W_{RAR} = \Gamma$, and thus;

$$K_{max} = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor \quad (43)$$

Generally, the number of RA slots falling within the backoff (BO) window is equal to the time of last sub-frame in the BO window minus the time before starting the BO window (divided by the interval between two consecutive RA slots), and it is given by:

$$N_{RA} = i + \left\lfloor \frac{T_{RAR} + W_{RAR} + W_{BO}}{T_{RA_REP}} \right\rfloor - i - \left\lfloor \frac{T_{RAR} + W_{RAR}}{T_{RA_REP}} \right\rfloor = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor \quad (44)$$

Depending on the values of Ψ and Γ , we have four cases:

1. Both Ψ/T_{RA_REP} and Γ/T_{RA_REP} are not integer values: in this case, we can write $\lceil \Psi/T_{RA_REP} \rceil$ as $(\lfloor \Psi/T_{RA_REP} \rfloor + 1)$ and $\lfloor \Gamma/T_{RA_REP} \rfloor$ as $(\lceil \Gamma/T_{RA_REP} \rceil - 1)$. Thus N_{RA} becomes:

$$N_{RA} = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor + 1 - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor + 1 = K_{max} + 2$$

where the value 2 represents the RA slots $x_a(i)$ and $x_d(i)$.

2. Ψ/T_{RA_REP} is integer whereas Γ/T_{RA_REP} is not: in this case, the value α_d (equation 11) is equal to zero and thus:

$$N_{RA} = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor + 1 = K_{max} + 1$$

where the value 1 represents the RA slot $x_a(i)$. Note that when x/y is an integer value, we have $x/y = \lceil x/y \rceil = \lfloor x/y \rfloor$.

3. Ψ/T_{RA_REP} is not integer whereas Γ/T_{RA_REP} is: in this case, the value α_a (equation 7) is equal to zero and thus:

$$N_{RA} = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor + 1 - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor = K_{max} + 1$$

where the value 1 represents the RA slot $x_d(i)$.

4. Both Ψ/T_{RA_REP} and Γ/T_{RA_REP} are integer values: in this case, the values α_a and α_d are equal to zero and thus:

$$N_{RA} = \left\lfloor \frac{\Psi}{T_{RA_REP}} \right\rfloor - \left\lfloor \frac{\Gamma}{T_{RA_REP}} \right\rfloor = K_{max}$$

Therefore, the equation giving K_{max} is true.

REFERENCES

- [1] R. Ratasuk, A. Prasad, Z. Li, A. Ghosh, and M. A. Uusitalo, "Recent advancements in M2M communications in 4G networks and evolution towards 5G," in *Proc. 18th Int. Conf. Intell. Next Gener. Netw. (ICIN)*, Feb. 2015, pp. 52–57.
- [2] L. Changwei, "Telco development trends and operator strategies," *WinWin Mag.*, vol. 13, pp. 19–22, Jul. 2012.
- [3] A. Zakrzewska, S. Ruepp, and M. Berger, "Towards converged 5G mobile networks-challenges and current trends," in *Proc. ITU Kaleidoscope Acad. Conf.: Living Converged World-Impossible Without Stand.?*, Jun. 2014, pp. 39–45.
- [4] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, and Y. Selén, "5G radio access," *Ericsson Rev.*, Jun. 2014.
- [5] 5GPPP. *The 5G Infrastructure Public-Private Partnership (5GPPP)*. 2013, [Online]. Available: <http://5g-ppp.eu/>
- [6] 3GPP R2-113198, "Further analysis of group paging for MTC," ITRI, May 2011.
- [7] C.-H. Wei, R.-G. Cheng, and S.-L. Tsao, "Performance analysis of group paging for machine-type-communications in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3371–3382, Sep. 2013.
- [8] C.-H. Wei, G. Bianchi, and R.-G. Cheng, "Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1940–1953, Apr. 2015.
- [9] O. Arouk, A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "On improving the group paging method for machine-type-communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 484–489.
- [10] R. Harwahyu, R.-G. Cheng, and R. F. Sari, "Consecutive group paging for LTE networks supporting machine-type communications services," in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 1619–1623.
- [11] R. Harwahyu, X. Wang, R. Sari, and R.-G. Cheng, "Analysis of group paging with pre-backoff," *EURASIP J. Wireless Commun. Net.*, vol. 2015, no. 1, pp. 1–9, 2015.
- [12] 3GPP R2-111916, "Backoff enhancements for RAN overload control," ZTE, Apr. 2011.
- [13] O. Arouk, A. Ksentini, and T. Taleb, "Group paging optimization for machine-type-communications," in *Proc. IEEE Ad-Hoc Sensor Netw. Symp. (ICC'15 (09) AHSN)*, London, U.K., Jun. 2015, pp. 6519–6524.
- [14] H. Thomsen, N. K. Pratas, E. Stefanovi, and P. Popovski, "Code-expanded radio access protocol for machine-to-machine communications," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 4, pp. 355–365, 2013.
- [15] 3GPP TR 37.868 V11.0.0, "Study on RAN improvement for machine-type-communications," Sep. 2012.
- [16] K.-D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Dec. 2011, pp. 373–377.

- [17] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: Overload control," *IEEE Netw.*, vol. 26, no. 6, pp. 54–60, Nov. 2012.
- [18] O. Arouk and A. Ksentini, "Multi-channel slotted aloha optimization for machine-type-communication," in *Proc. 17th ACM Int. Conf. Model. Anal. Simul. Wireless Mobile Syst.*, 2014, pp. 119–125.
- [19] D. Wiriaatmadja and K. W. Choi, "Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 33–46, Jan. 2015.
- [20] 3GPP TS 23.682 V11.4.0, "Architecture enhancements to facilitate communications with packet data networks and applications," Jun. 2013.
- [21] 3GPP TS 29.368 V12.2.0, "TSP interface protocol between the MTC inter working function (MTC-IWF) and service capability server (SCS)," 2014.
- [22] 3GPP TR 36.912 V11.0.0, "Feasibility study for further advancements for E-UTRA (LTE-Advanced)," Sep. 2012.
- [23] 3GPP TR 36.300 V12.3.0, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); Overall description," Sep. 2014.
- [24] L. Sanguinetti and M. Morelli, "An initial ranging scheme for the IEEE 802.16 OFDMA uplink," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3204–3215, Sep. 2012.
- [25] C.-L. Lin and S.-L. Su, "A robust ranging detection with MAI cancellation for OFDMA systems," in *Proc. 13th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2011, pp. 937–941.
- [26] H. Minn and X. Fu, "A new ranging method for OFDMA systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'05)*, Nov. 2005, vol. 3, p. 6.
- [27] L. Sanguinetti, M. Morelli, and L. Marchetti, "A random access algorithm for LTE systems," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 1, pp. 49–58, 2013.
- [28] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 180–185, Jan. 2015.
- [29] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmwave communications," in *Proc. 48th Asilomar Conf. Signals Syst. Comput.*, Nov. 2014, pp. 1926–1930.
- [30] 3GPP TS 36.213 V12.4.0, "Physical layer procedures," Dec. 2014.
- [31] 3GPP TS 36.101 V12.6.0, "User equipment (UE) radio transmission and reception," Dec. 2014.
- [32] 3GPP TS 36.321 V12.4.0, "Medium access control (MAC) protocol specification," Dec. 2014.
- [33] 3GPP TS 36.331 V12.4.1, "Radio resource control (RRC); Protocol specification," Dec. 2014.
- [34] H. L. Bertoni, *Radio Propagation for Modern Wireless Systems*. Englewood Cliffs, NJ, USA: Prentice Hall, 1999.
- [35] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Pearson Education, 2007.
- [36] Wikipedia. *Exponential Function* [Online]. Available: http://en.wikipedia.org/wiki/Exponential_function
- [37] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, "Overload control for machine-type-communications in LTE-advanced system," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 38–45, Jun. 2012.



Osama Arouk (S'14) received the B.S. degree in telecommunications engineering from Aleppo University, Aleppo, Syria, in 2009, and the M.S. degree in signal processing from the University of Rennes 1, Rennes, France, in 2012. He is currently pursuing the Ph.D. degree at INRIA/IRISA, Rennes, France. His research interest includes machine-type-communication (MTC), Internet-of-Things (IoT), performance evaluation, multichannel slotted aloha systems, performance analysis of LTE, and beyond networks.



Adlen Ksentini (SM'14) received the M.Sc. degree in telecommunication and multimedia networking from the University of Versailles Saint-Quentin-en-Yvelines, Versailles, France, and the Ph.D. degree in computer science from the University of Cergy-Pontoise, Cergy-Pontoise, France (with a dissertation on QoS provisioning in the IEEE 802.11-based networks), in 2005. He is currently an Associate Professor with the University of Rennes 1, Rennes, France. He is a member of the Dionysos Team with INRIA, Rennes, France. He is involved in several national and European projects on QoS and QoE support in future wireless and mobile networks. He has coauthored over 80 technical journal and international conference papers. His research interests include future Internet networks, mobile networks, QoS, QoE, performance evaluation, and multimedia transmission. He is the TPC Chair of the Wireless and Mobile (WMN) Symposium of the IEEE ICC 2016. He was a Guest Editor of the *IEEE Wireless Communication Magazine*, *IEEE Communication Magazine*, and two ComSoc MMTS Letters. He has been on the technical program committee of major IEEE ComSoc, ICC/Globecom, ICME, WCNC, and PIMRC conferences. He was the recipient of the Best Paper Award from the IEEE ICC 2012 and ACM MSWiM 2005.



Tarik Taleb (S'05–M'05–SM'10) received the B.E. degree (with distinction) in information engineering, and the M.Sc. and Ph.D. degrees in information sciences from the Graduate School of Information Sciences (GSIS), Tohoku University, Sendai, Japan, in 2001, 2003, and 2005, respectively. He is currently a Professor with the School of Electrical Engineering, Aalto University, Espoo, Finland. Before that, he was a Senior Researcher and 3GPP Standards Expert with NEC Europe Ltd, Heidelberg, Germany. He was then leading the NEC Europe Labs Team working on RD projects on carrier cloud platforms, an important vision of 5G systems. Before joining NEC and until March 2009, he was an Assistant Professor with the GSIS, Tohoku University, in a laboratory fully funded by KDDI, the second largest network operator in Japan. From October 2005 to March 2006, he was a Research Fellow with the Intelligent Cosmos Research Institute, Sendai, Japan. His research interests include architectural enhancements to mobile core networks (particularly 3GPPs), mobile cloud networking, network function virtualization, software defined networking, mobile multimedia streaming, inter-vehicular communications, and social media networking. He has been also directly engaged in the development and standardization of the Evolved Packet System as a member of 3GPPs System Architecture working group. As an attempt to bridge the gap between academia and industry, he founded the IEEE Workshop on Telecommunications Standards: From Research to Standards, a successful event that was recognized with the Best Workshop Award by the IEEE Communication Society (ComSoC). Based on the success of this workshop, he has also founded and has been the Steering Committee Chair of the IEEE Conference on Standards for Communications and Networking. He is the General Chair of the 2019 edition of the IEEE Wireless Communications and Networking Conference (WCNC19) to be held in Marrakech, Morocco. He is/was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *IEEE Wireless Communications Magazine*, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and a number of Wiley journals. He is serving as the Chair of the Wireless Communications Technical Committee, the largest in the IEEE ComSoC. He also served as the Vice Chair of the Satellite and Space Communications Technical Committee of the IEEE ComSoc (2006–2010). He has been on the technical program committee of different IEEE conferences, including Globecom, ICC and WCNC, and chaired some of their symposia. He was the recipient of the 2009 IEEE ComSoc Asia-Pacific Best Young Researcher Award (June 2009), the 2008 TELECOM System Technology Award from the Telecommunications Advancement Foundation (March 2008), the 2007 Funai Foundation Science Promotion Award (April 2007), the 2006 IEEE Computer Society Japan Chapter Young Author Award (December 2006), the Niwa Yasujirou Memorial Award (February 2005), and the Young Researcher's Encouragement Award from the Japan chapter of the IEEE Vehicular Technology Society (VTS) (October 2003). Some of his research work have been also awarded best paper awards at prestigious conferences.