Abstract—This letter considers a wireless M2M communication scenario with a massive number of M2M devices. Each device needs to send its reports within a given deadline and with certain reliability, e. g. 99.999%. A pool of resources available to all M2M devices is periodically available for transmission. The number of transmissions required by an M2M device within the pool is random due to two reasons - random number of arrived reports since the last reporting opportunity and requests for retransmission due to random channel errors. We show how to dimension the pool of M2M-dedicated resources in order to guarantee the desired reliability of the report delivery within the deadline. The fact that the pool of resources is used by a massive number of devices allows to base the dimensioning on the central limit theorem. The results are interpreted in the context of LTE, but they are applicable to any M2M communication system.

Index Terms—Wireless cellular access, M2M communications, LTE

I. INTRODUCTION

In the past two decades, the main focus of the cellular access engineering was on the efficient support of human-oriented services, like voice calls, messaging, web browsing and video streaming services. A common feature of these services is seen in the relatively low number of simultaneous connections that require high data rates. On the other hand, the recent rise of M2M communications introduced a paradigm shift and brought into research focus fundamentally new challenges. Particularly, M2M communications involve a massive number of low-rate connections, which is a rather new operating mode, not originally considered in the cellular radio access.

Smart metering is a showcase M2M application consisting of a massive number of devices, up to 30000 [1], where meters periodically report energy consumption to a remote server for control and billing purposes. On the other hand, the report size is small, of the order of 100 bytes [2]. The current cellular access mechanisms, considering all the associated overhead, can not support this kind of operation [3]. There are on-going efforts in 3GPP that deal with the cellular access limitations, investigating methods for decreasing the access overhead for small data transmissions [4], access overload control [5] and guaranteed quality of service [6]. Besides LTE, [7] proposes an allocation method for reports with deadlines in GPRS/EDGE, showing that up to 10^4 devices can be effectively supported in a single cell by avoiding random access and using a periodic structure to serve the devices such that the deadlines are met.

In this letter, we consider a system with a periodically occurring pool of resources that are reserved M2M communications and shared for uplink transmission by all M2M devices. The reoccurring period is selected such that if a report is transmitted successfully within the upcoming resource pool, then the reporting deadline is met. We note that, if each device has a deterministic number of packets to transmit in each resource pool and if there are no packet errors, then the problem is trivial, because a fixed number of resources can be pre-allocated periodically to each device. However, if the number of packets, accumulated between two reporting instances, is random and the probability of packet error is not zero, then the number of transmission resources required per device in each transmission period is random. On the other hand, as the number of transmission resources in each instance of the resource pool is fixed, the following question arises: How many periodically reporting devices can be supported with a desired reliability of report delivery (i.e., 99.999%), for a given number of resources reserved for M2M communications? We elaborate and analyze the proposed approach in LTE context; however, the presented ideas are generic and implementable in other systems where many devices report to a single base station or access point.

The rest of the letter is organized as follows. Section II presents the system model. Section III is devoted to the analysis of the proposed access method. Section IV presents the numerical results and Section V concludes the letter.

II. SYSTEM MODEL

We focus on the case of periodic reporting, where the length of the reporting interval (RI), denoted by \( T_{RI} \), depends on the application requirements [2]. The M2M resources for uplink transmission are reserved to occur periodically, at the end of each RI. The periodic reporting is typically modeled either as a Poisson process with arrival rate \( \lambda = 1/T_{RI} \), where devices can actually send none, one, or multiple reports within RI, or as a uniform distribution, where devices send a single packet per RI [3], [2]. Our analysis will focus on the former case, but we note that the derived results can be readily applied to the latter arrival model, as well. We assume that all report arrivals that occur within the current reporting interval are served in the next reporting interval.

The LTE access combines TDMA and FDMA, such that access resources are represented in 2D, see Fig. I. As depicted, time is organized in frames, where each frame is composed of...
subframes with duration $T_s = 1$ ms. The minimum amount of uplink resources that can be allocated to a device is one resource block (RB), corresponding to a single subframe and 12-subcarriers in frequency. We assume that the uplink resources are split into two pools, one reserved for M2M services (depicted in blue in Fig. 1), and the other used for other services. Note that the approach of splitting the resources for M2M and non-M2M has often been used [10], [11], as their requirements are fundamentally different. Finally, we assume that there is a set of $Y$ RBs reserved for M2M resource pool in each subframe.

The M2M resource pool is divided into two parts, denoted as the preallocated and common pool, which reoccur with period $T_{RI}$, as depicted in Fig. 2a). We assume that there are $N$ reporting devices, and each device is preallocated an amount of RBs from the preallocated pool dimensioned to accommodate a single report and an indication if there are more reports, termed excess reports, from the same device to be transmitted within the same interval. The common pool is used to allocate resources for the excess reports, as well as all the retransmissions of the reports/packets that were erroneously received. These resources can only be reactively allocated and in our case we consider the LTE data transmission scheme, where each transmission has an associated feedback that can be used to allocate the resources from the common pool.

The length of the M2M resource pool, preallocated pool and common pool, expressed in number of subframes, are denoted by $X$, $X_P$ and $X_C$, respectively, see Fig. 2b), such that

$$X = X_P + X_C = \alpha N + X_C.$$  

(1)

where $\alpha \leq 1$ denotes the fraction of RBs per subframe required to accommodate a report transmission and where the value of $X_C$ should be chosen such that a report is served with a required reliability. The analysis how to determine $X_C$, given the constraints of the required number of RBs per report, number of devices and reliability, is the pivotal contribution of this paper given the constraints of the required number of RBs per report, as well as all the retransmissions of the reports/packets that were erroneously received. The straightforward way to characterize $R$ is to derive its pmf.

Denote by $W_{i,j}$ a random variable that models the total number of transmissions of report $j$ from device $i$; i.e., $W_{i,j}$ includes the first transmission and any subsequent retransmissions that may occur due to reception failures. Further, assume that the maximum value of $W_{i,j}$ is limited to $L$, where $L$ is a system parameter, applicable to all reports from all devices. The probability mass function (pmf) of $W_{i,j}$ can be modeled as a geometric distribution truncated at $L$:

$$P[W_{i,j} = k] = \begin{cases} \frac{p_{e}^{k-1}(1-p_{e})}{p_{e}^{k-1}}, & 1 \leq k \leq L - 1, \\ 1, & k = L, \end{cases}$$  

(2)

where $p_{e}$ is the probability of a reception failure.

Recall that the reporting of device $i$ is modeled as a Poisson process $U_i$ with arrival rate $\lambda$, where the device can send none ($U_i = 0$), one ($U_i = 1$) or multiple reports ($U_i \geq 1$) within $T_{RI}$. As stated above, the first transmission of the first report of a device is sent in the preallocated pool, while all subsequent transmissions take place in the common pool. These include: a) potential retransmissions of the first report, and b) transmissions and potential retransmissions of all excess reports. Denote by $R_i$ the random variable that corresponds to the total number of transmissions from device $i$ requiring resources in the common pool:

$$R_i = \begin{cases} 0, & U_i = 0, \\ \sum_{j=1}^{U_i} W_{i,j} - 1, & U_i \geq 1. \end{cases}$$  

(3)

The total number of transmissions of all devices requiring resources from the common pool is:

$$R = \sum_{i=1}^{N} R_i.$$  

(4)

The straightforward way to characterize $R$ is to derive its pmf. However, as the supposed number of reporting devices is very large, $R$ is a sum of a large number of independent identically distributed (IID) random variables. Therefore, we apply the central limit theorem and assume that $R$ is a Gaussian random variable, requiring characterization only in terms of

![Fig. 1. Representation of the LTE uplink resource structure, where a set of RBs has been reserved for M2M purposes.](image1)

![Fig. 2. a) Periodically occurring M2M resource pool. b) Division of M2M resource pool in the pre-allocated and common pool.](image2)
the expectation and variance. We proceed by evaluation of \( E[R] \) and \( \sigma^2[R] \), and show in Section IV that this approach provides accurate results.

**A. Expectation and variance of \( R \)**

The expectation of \( R \) is:

\[
E[R] = E \left[ \sum_{i=1}^{N} R_i \right] = N \cdot E[R_i]. \tag{5}
\]

Taking into account (3), it could be shown that:

\[
E[R_i] = E[R_i|U_i=0]P[U_i=0] + E[R_i|U_i \geq 1]P[U_i \geq 1],
= E[R_i|U_i \geq 1]P[U_i \geq 1],
= E \left[ \sum_{j=0}^{W_i} W_{i,j} - 1 \right| U_i \geq 1] \left( 1 - e^{-\lambda_T R_i} \right). \tag{6}
\]

By putting \( \lambda T R_i = 1 \) and applying Wald’s equation (12), the equation (6) becomes:

\[
E[R_i] = (E[U_i|U_i \geq 1]E[W_{i,j} - 1] - 1) \left( 1 - e^{-1} \right),
= \frac{1 - p_e^L}{1 - p_e} - (1 - e^{-1}). \tag{7}
\]

where we used the fact that \( E[U_i|U_i \geq 1] = 1/(1 - e^{-1}) \).

Substituting (7) into (5) yields expectation of \( R \).

The variance of \( R \) can be derived in a similar fashion, using the identities related to the variance of the random sum of random variables (12). Due to the space limitations, we omit the derivation and present only the final result:

\[
\sigma^2[R] = N \left[ \frac{(2L-1)p_e^{L+1} - (2L+1)p_e^L + p_e + 1}{(1 - p_e)^2}
+ e^{-1} \cdot \left( 1 - 2 \cdot \frac{1 - p_e^L}{1 - p_e} - e^{-1} \right) \right]. \tag{8}
\]

**B. Probability of Report Failure**

In this subsection, we assess the probability of report failure, i.e., the probability that the report has not been successfully delivered after all attempted (re)transmissions. In general, this probability depends both on the number of resources available and the scheduling policy applied in the common pool. In order to avoid the particularities related to scheduling, we derive an upper bound on the probability of failure which is valid for any scheduling policy.

Denote by \( \Phi \) the event that a report experiences a failure. Further, denote by \( l \) the number of required report transmissions, which includes the first transmission and all the required retransmissions. If we assume for a moment that the number of available resources in the common pool is infinite, then the report fails to be delivered only when the required number of transmissions exceeds \( L \):

\[
P_{\infty}[\Phi] = P_{\infty}[\Phi, l > L] = P[l > L] = p_e^L. \tag{9}
\]

If the common pool that consists of \( X_C \) subframes can accommodate \( C \) transmissions (i.e., \( C \) is the capacity of the common pool in number of transmissions), then the probability of report failure can be written as:

\[
P[\Phi] = \sum_{k=1}^{L} P[\Phi, l = k] + P[\Phi, l > L],
= \sum_{k=1}^{L} P[\Phi|l = k]P[l = k] + p_e^L. \tag{10}
\]

Further, for \( 1 \leq k \leq L \):

\[
P[\Phi|l = k] = P[\Phi, R > C|l = k] + P[\Phi, R \leq C|l = k],
= P[R > C|P[\Phi|l = k, R > C], \tag{11}
\]

where we used the fact that \( P[\Phi, R \leq C|l = k] = 0 \), i.e., there is no report delivery failure when the total number of required transmissions \( R \) is not greater than the capacity of the common pool \( C \), for \( l \leq k \leq L \). Regardless of the scheduling policy applied in the common pool, it is always \( P[\Phi|l = k, R > C] \leq 1 \), which leads to the following upper bound:

\[
P[\Phi|l = k] \leq P[R > C]. \tag{12}
\]

Finally, substituting (12) into (10) yields:

\[
P[\Phi] \leq Q \left( \frac{C - \mu}{\sigma} \right) + \eta E[R] \tag{13}
\]

where \( \mu = E[R], \sigma = \sqrt{\sigma^2[R]} \), and \( Q(\cdot) \) is Q-function, standardly used for Gaussian random variables.

**IV. RESULTS**

We first validate the Gaussian assumption in the analysis by comparing the probability density function (pdf) and cumulative density function (cdf) of the Gaussian model with the simulated ones. Fig. 3 presents a tight match between the model and simulations, when number of reporting devices is \( N = 100 \), the maximum number of transmissions per report is \( L = 10 \), the probability of report error \( p_e \) takes values 0.1 and 0.4, respectively, and the number of the simulation runs is set to \( 10^5 \) for each parameter combination.

Further, using the bound derived in (13), we determine the percentage of LTE system resources required for reliable M2M services, defined as the ratio of RBs required for M2M and the total amount of RBs available in the system (see Fig. 1). LTE foresees a maximum bandwidth of 100 RBs (i.e., 20 MHz) per subframe, see Fig. 1 while 20 RBs per subframe (i.e., 5 MHz) are most commonly used [13]. The amount of RBs required to
the report size has a large impact in the results, demanding up
64-QAM) LTE modulation schemes. requests for retransmission due to random channel errors. The
of arrived reports since the last reporting opportunity and (2)
the pool is random due to two reasons: (1) random number
number of transmissions required by an M2M device within
order that relies on a pool of resources reoccurring
64
modulation scheme is
1
report transmission depends on the used modulation
1
report size (RS). For the individual transmission error we used
p_r = 0.1, which is a typical value for LTE data transmissions
[14], while the maximum number of report transmissions is
again set to L = 10. The maximal number of devices is set
to 30000, as indicated by 3GPP in [1]. Finally, we set the
probability of report failure to P[Φ] ≤ 10^{-3}, i.e. the desired
reliability to at least 99.99%. To validate the analytical upper
bound, we simulated a random scheduler with 10^7 repeats for
each parameter combination.

Fig. 4 shows the percentage of system resources required
to serve N devices for a reporting interval RI of 1 minute,
which corresponds to the most demanding RI according to
[2]. The report size (RS) is 100 bytes. It can be seen that
in the worst case, for the available bandwidth of 5 MHz and
the lowest-order modulation (QPSK), up to 30000 devices
can be served with only 9% of the available system resources.
If a larger bandwidth (5 MHz) and/or higher-order modulations
(64-QAM) is used, then only 3% of the available resources
are required to achieve the target report reliability.

Fig. 5 depicts the required fraction of system capacity for
M2M service, when the report RS varies between 100 bytes
and 1 kbytes [2], the system bandwidth is set to 5 MHz,
modulation scheme is 64-QAM, and p_r = 0.1. Obviously,
the report size has a large impact in the results, demanding up
30% of the system capacity in the worst case.

Finally, we note that Figs. 4 and 5 also demonstrate a tight
match between the analytical and simulated results.

V. CONCLUSIONS

We have introduced a contention-free allocation method
for M2M that relies on a pool of resources reoccurring
periodically in time. Within each occurrence, feedback is used
to reactively allocate resources to each individual device.
The number of transmissions required by an M2M device within
the pool is random due to two reasons: (1) random number
of arrived reports since the last reporting opportunity and (2)
requests for retransmission due to random channel errors. The

\[ P[Φ] \leq 10^{-3}, \text{ RI of 1 minute, bandwidth of 5 MHz, 64-QAM and } p_r = 10^{-1}. \]

objective is to dimension the pool of M2M-dedicated resources
in order to guarantee certain reliability in the delivery of a
report within the deadline. The fact that the pool of resources
is used by a massive number of devices allows to base the
dimensioning on the central limit theorem. Promising results
have been shown in the context of LTE, where even with
the lowest-order modulation only 9% of the system resources
are required to serve 30K M2M devices with a reliability of
99.99% for a report size of 100 bytes. The proposed method
can be applied to other systems, such as 802.11ah.

REFERENCES

[1] 3rd Generation Partnership Project (3GPP), “RACH Intensity of Time
Controlled Devices,” TSG RAN WG2 R2-102340, 2010.
[2] 3rd Generation Partnership Project (3GPP), “Smart Grid Traffic Be-
haviour Discussion,” TR R2-102340, 2010.
[3] A. Laya, L. Alonso, and J. Alonso-Zarate, “Is the Random Access
Channel of LTE and LTE-A Suitable for M2M Communications? A
Survey of Alternatives,” IEEE Commun. Surveys Tuts., vol. 16, no. 1,
pp. 4–16, First 2014.
[4] 3rd Generation Partnership Project (3GPP), “Study on Machine-
Type Communications (MTC) and other mobile data applications
communications enhancements,” TR 22.368, Dec. 2013.
[5] 3rd Generation Partnership Project (3GPP), “System improvements for
Machine-Type Communications (MTC),” TR 23.888, 2013.
[6] S.-Y. Lien and K.-C. Chen, “Massive Access Management for QoS Guar-
antees in 3GPP Machine-to-Machine Communications,” IEEE Comm.
Lett., vol. 15, no. 3, pp. 311–313, May 2011.
[7] A. Laya, L. Alonso, J. Alonso-Zarate, and R. Almendros, “How Many
Smart Meters can be Deployed in a GSM Cell?” in IEEE ICC’13
Workshop Telecom R25, Budapest, Hungary, Jun. 2013.
[8] 802.16 Broadband Wireless Access Working Group (802.16p),
“IEEE 802.16d Machine to Machine (M2M) Evaluation Methodology
Document (EMD),” EMD 11/0014, May 2011.
[9] 3rd Generation Partnership Project (3GPP), “USF Capacity Evaluation
for MTC,” TSG GERAN #46 GP-100984, 2010.
[10] K.-D. Lee, S. Kim, and B. Yi, “Throughput Comparison of Random
Access Methods for M2M Service over LTE networks,” in IEEE GC’11
Workshops, Dec. 2011.
[11] R. Posor, H. Wilhelm, M. Sadykov, D. Navratil, and T. Pape, “Overload
Control Method for Synchronized MTC Traffic in Geran,” in IEEE VTC
Fall 2011, Sep. 2011.
[12] S. M. Ross, Introduction to probability and statistics for engineers and
scientists. Academic Press, 2004.
[13] B. Yang, G. Zhu, W. Wu, and Y. Gao, “M2M Access Performance in
LTE-A System,” Transactions on Emerging Technologies, vol. 25, no. 1,
pp. 3–10, 2014.
[14] A. Ahmad, LTE-Advanced: A Practical Systems Approach to Under-
standing 3GPP LTE Releases 10 and 11 Radio Access Technologies.
Elsevier Science, 2013.