How Many Smart Meters can be Deployed in a GSM cell?

Madueño, Germán Corrales; Stefanovic, Cedomir; Popovski, Petar

Published in:
Proceedings of IEEE International Conference on Communications (ICC 2013). IEEE, 2013.

DOI (link to publication from Publisher):
10.1109/ICCW.2013.6649431

Publication date:
2013

Document Version
Early version, also known as pre-print

How Many Smart Meters can be Deployed in a GSM cell?

Germán Corrales Madueño, Čedomir Stefanović, Petar Popovski

Department of Electronic Systems, Aalborg University, Denmark
Email: {gco,cs,petarp}@es.aau.dk

***Abstract*—The need to deploy large number of wireless devices, such as electricity or water meters, is becoming a key challenge for any utility. Furthermore, such a deployment should be functional for more than a decade. Many cellular operators consider LTE to be the single long term solution for wide area connectivity serving all types of wireless traffic. On the other hand, GSM is a well-adopted technology and represents a valuable asset to build M2M infrastructure due to the good coverage, device maturity, and low cost. In this paper we assess the potential of GSM/GPRS/EDGE to operate as a dedicated network for M2M communications. In order to enable M2M-dedicated operation in the near future, we reengineer the GSM/GPRS/EDGE protocol in a way that requires only minor software updates of the protocol stack. We propose different schemes to boost the number of M2M devices in the system without affecting the network stability. We show that a single GSM cell can support simultaneous low-data rate connections (e. g. to smart meters) in the order of 10^4 devices.**

I. Introduction

In the emerging area of machine-to-machine (M2M) communications, smart metering is a showcase application: a large number of electricity/water/heat meters use sophisticated wireless networking for two-way communication with a central controller/data collector. The usage of wireless techniques for M2M communication has been made possible due to the level of maturity attained by the wireless technologies: small, inexpensive embedded devices have significant computational power and operate at very low power levels.

M2M communication has significantly different requirements from, e. g. human-to-machine (H2M) services (download, web browsing, video streaming), where large data volumes are sent and high data rate is required. In majority of the scenarios, M2M communication is based on intermittent transmission/reception of small data portions and pose requirements that are different from the ones according to which the common wireless protocols are designed. Some of the most important requirements are the following:

- Transmission from a massive number of devices and maintenance of a large number of active connections;
- Ability to send a small amount of data while decreasing the overhead percentage;
- Real-time communication with low latency;
- Certain connections that carry critical control data require a high degree of reliability, such that a connection should be kept alive more than 99.XX % of the time.

These requirements become more challenging when one considers the forecasts that state that by 2020 there will be 50 billion M2M connected wireless devices [1], spanning a wide application range: smart grid, smart metering, control/monitoring of homes and industry, e-health, etc. While there are many ongoing standardization activities [2], M2M communication solutions have started to be deployed through the existing cellular interfaces, such as GSM and LTE. Specifically, there is an increasing deployment of cellular-based wireless smart meters, such as the ones based on GPRS [3]. Some of those deployments are very large, such as Hydro-Quebec in Canada [4], with about 3.8 million devices that periodically send only a few bytes (KW/h consumption for instance). Another example is happening in Spain and Portugal, where Endesa, the largest Iberian operator, will replace a total of 13 million electric meters with smart meters by 2018 [5]. Since neither GSM nor LTE are originally designed to support massive M2M communication, there are ongoing research and standardization activities to modify those interfaces, notably LTE, in order to support the M2M traffic characteristics [6].

In this paper we explore the potential of GSM to operate as an M2M-dedicated network and support a large number of active connections. The advantages of GSM as a basis for a future M2M infrastructure include ubiquitous coverage, worldwide frequency availability, device maturity, and low cost. In our approach, the GSM[1] protocol is reengineered in a way that the physical layer is kept intact, the frame sent over the air has a backward-compatible structure, while changes are made in the algorithms and the local protocol-related variables that are used by the Base Station and the M2M devices. The key idea is to enable the communication nodes to address/use the resources sent over the air with a finer level of granularity. Such an approach allows smooth evolution of the current 2G networks towards M2M-dedicated networks. We have analyzed and evaluated the capacity of the reengineered GSM network to support the M2M traffic patterns prescribed by 3GPP [6]. The evaluation is made for a traffic with limited latency, such that if a packet is in a queue for excessive time, it is dropped. Our initial results are very promising, confirming that it is possible to scale GSM towards a regime in which a single cell and over a signal

[1]Hereinafter, we use the generic term GSM to refer to data services on GSM, GPRS, and EDGE.

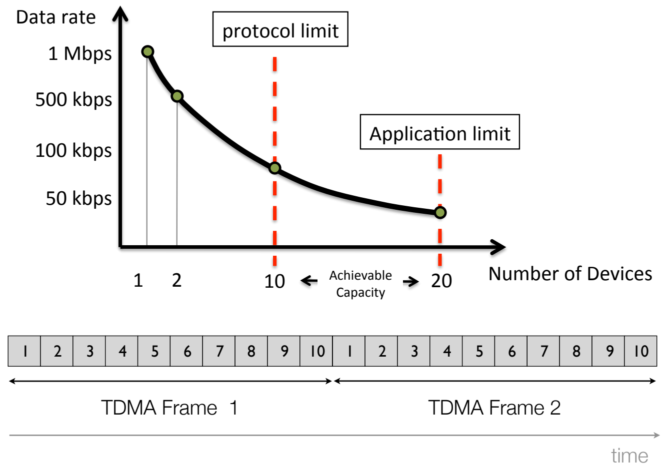

Fig. 1: Ideal system in which the bandwidth is shared among the multiplexed devices. The protocol operation is limiting the number of devices, despite the application requirements.

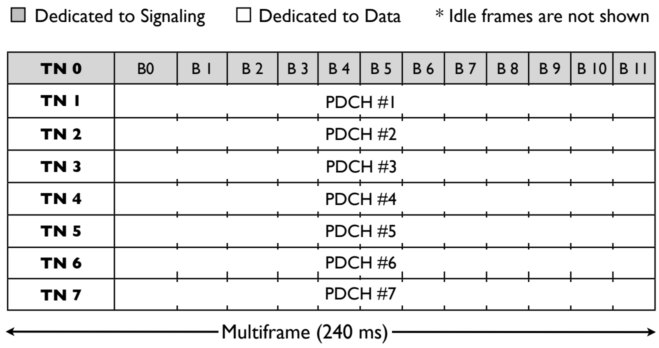

Fig. 2: Multiframe structure composed by 7 time slots (TN0 to TN7) and 7 Packet Data Channels (PDCH).

frequency channel, the number of active low-rate connections is very large ($> 10^4$). Moreover, we show that our system can achieve a good tradeoff, in a large range, between the packet dropping probability and the total number of meters served in a cell.

The rest of the paper is organized as follows. Section II presents the relevant aspects of GSM that have influence on the analysis and the major limitations in regards to M2M. Section III, gives in a compressed manner what is to be modified in the current GSM system. In Section IV, we present the mathematical model used to characterize the system. In Section V the tradeoff between the number of meters and the dropping probability that can be achieved in a GSM system is presented. The last section provides a conclusion.

II. Protocol Limitations for Massive M2M Communication

In this section we first discuss the generic case of TDMA system, and illustrate the mismatch between application requirements and protocol design. Then, we briefly describe the related limitations of the GSM radio access.

A. Protocol Limit in a Generic TDMA System

Ideally, a TDMA system should be able to allocate as many as possible devices as long as the quality of service is guaranteed. The target operation can be described as follow: if a nominal data rate of a system is R bps and there are N potential users, in the ideal case the bandwidth is shared such that each device gets a data rate of R/N bps. At the same time, the limit on N should be posed by the application requirements, rather than protocol limitations. However, in practice, systems are typically not able to operate in this manner.

For example, consider an example TDMA system in which a time frame consists of 10 slots. Each slot has a duration of 1 ms and it carries 10 Kbits; the total available bandwidth is thus 1 Mbps. Further, assume that a device can be allocated from one to all ten slots, while the minimum bandwidth needed by a device to properly operate is 50 Kbps. Fig. 1 illustrates how the system operates: when the system is empty and a new device arrives, the full bandwidth (i.e., 1 Mbps) is allocated to the only user. Each time a new device arrives, the system re-allocates the slots so that all devices get a data rate of $1/N$ Mbps, where N is current number of devices admitted in the system. A data rate of 100 Kbps per device is provided when the system if full, when there are 10 devices being served. Due to *protocol limitation, i.e., the granularity of available resources*, any new arrival is rejected, even though the bandwidth is enough to serve up to 20 devices.

B. GSM Protocol and its Limitations for M2M

The GSM radio-link is TDMA based, where both uplink and downlink are organized in *multiframes*. A multiframe has duration of 240 ms and is composed of 12 radio blocks and 8 time slots. Its structure is depicted in Fig. 2. A radio block is the minimum amount of information that can be sent or received by the devices. Time slots are independent TDMA channels, usually denoted as Packet Data Channels (PDCHs) that can be used for signaling or data-transmission; in this paper we assume a typical configuration, where PDCH #0 is dedicated to signaling, while the remaining seven PDCHs carry data.

The phases of the radio access in GSM are resource request, data transmission/reception and resource release. The resource-request procedure is as follows. First, a device accesses the medium by sending a request message in a so-called Packet Random Access Channel (PRACH), which is a logical channel defined over a chosen PDCH. Upon reception of the PRACH request, the base station grants the resource request assigning the resources in one or more PDCHs. Alternatively, the base station can reject a request if there are no free resources. The grant message is carried in another logical channel - the Packet Access Grant Channel (PAGCH). This message contains the PDCH and the Uplink Status Flag (USF) allocated to the device. Finally, the device is allowed to transmit only in block $k+1$ of the allocated PDCH in the uplink if its USF was announced in block k of the same PDCH in the downlink.

Through the above stages of a random access procedure, 3GPP recommends that a blocking probability of 2% per stage should not be exceeded [7]. Furthermore, a numerical comparison of the bottlenecks in the respective stages of the procedure is presented in [7], where the USF constraint happens to be the most limiting factor. Particularly, as a USF is 3 bits long and the value 000 is reserved, a maximum of 7 devices can be multiplexed per PDCH, resulting in a maximum of 49 simultaneously allocated devices (i.e, 7 USFs in 7 data PDCHs)[2].

III. Reengineering the GSM system

In this section we propose the solution to overcome the USF limitations by allocating/multiplexing more devices per PDCH. The allocation space is expanded by reinterpretation the usage of USF.

The main idea is that a USF value does not hold for all the blocks of the allocated PDCHs anymore, but is valid only for a subset of blocks, during a predefined set of reoccurring multiframes. Hence, several devices could be identified with the same USF in the same PDCH and block, but the method is collision free as devices are allocated in different multiframe sets, and thus orthogonally in time.

The flow of the procedure is represented in Fig. 3. The allocator for the device now consists of the following parameters: PDCHs, USF, K and M, where K is the number of blocks per PDCH for which the USF is valid and M is the period by which the allocated multiframes reoccur. We assume that M and K are equal for all devices (i.e., we consider a case with homogeneous traffic), these parameters can be broadcast in the Cell Information messages. The device is only allowed to transmit if the USF received in the downlink matches its own in the specific combination of multiframes, PDCHs and blocks; this approach allows for much finer granularity than in the ordinary GSM case.

For a better understanding we provide an example (see Fig. 4) in which 4 devices are multiplexed into a single PDCH with 3 USFs. The parameter M is set to two, therefore, devices have a chance to transmit every second multiframe. Devices #1, #2 have been granted access in multiframe 1 in block 0 with $USF1$, $USF2$ respectively. They all transmit without collisions because they have been assigned to different USF values (this is an example of the ordinary GSM multiplexing strategy). Then, in multiframe 4, device #3 gets the same USF value than device #1 in the same block 0. No collision occurs because #1 transmits in even multiframes, while #3 transmits only in odd multiframes. Finally, in multiframe 7, a new device, device #4, arrives, for which the base station allocates block 1 with USF3, again, no collision occurs as both devices are assigned to independent channels (block 0 and block 1).

[2]We assume that only one radio channel is used. In general case, if there are B radio channels, then a maximum of $49B$ devices can be admitted in the system.

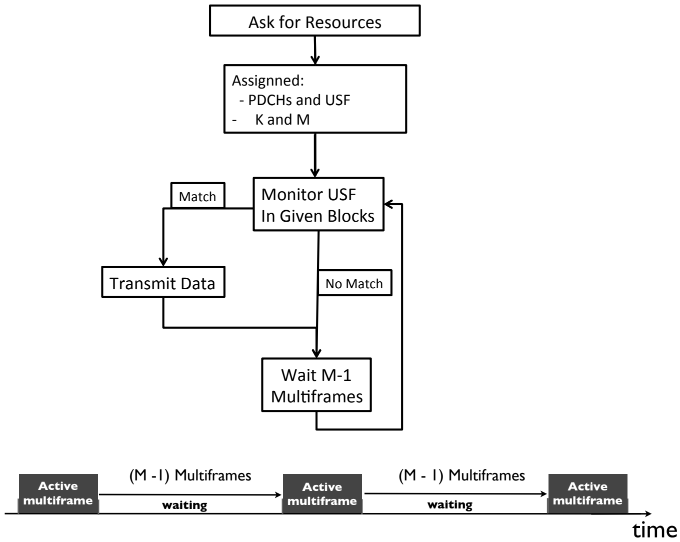

Fig. 3: Flow-chart outlining the proposed method.

IV. Analysis

In this section we first describe a system that can host a large number of smart meters at a target data rate. Moreover, we introduce the requirements of smart meters. Subsequently, we present the maximum achievable capacity in the system for a given amount of resources. Later, we introduce the analytical tools to analyze the performance in terms of reliability.

A. System Model

The system can be described as follows: first the device requests resources, indicating the amount of data needed and the reporting interval RI. Then the base station grants access and schedules the device to transmit in specific time intervals, allowing an efficient sharing of the resources among all the devices in the cell in a coordinated way. The system essentially operates in a circuit-switched manner, where each meter is allocated a portion of the link time. The circuit-switched structure is periodically reestablished (e. g., once a day, week, month, etc.) through the PRACH procedure. The resources used for PRACH are considered negligible.

Smart meters are expected to be installed in fixed locations with high device intensity, where the traffic patterns correspond to device-originated transmissions with small payload (in the range of 100 to 1000 bytes) and periodical reporting in 1 min, 5 mins, 15 mins, 1 hour, 6 hours, 12 hours and 24 hours [8]. In addition, devices tolerate a delay up to the next scheduled transmission opportunity if the message was not successfully delivered. Furthermore, a delay of up to 1 min is allowed in event-triggered alarms [8].

In this paper, we consider a delay tolerance equal to the report interval; any message arriving beyond the delay tolerance (deadline) is dropped and counted as a lost report. The loss probability is defined as the long-run fraction of messages that do not arrive after the deadline.

We model the traffic by a Poisson process as defined by 3GPP [9] [10] and IEEE 802.16p Machine to Machine group

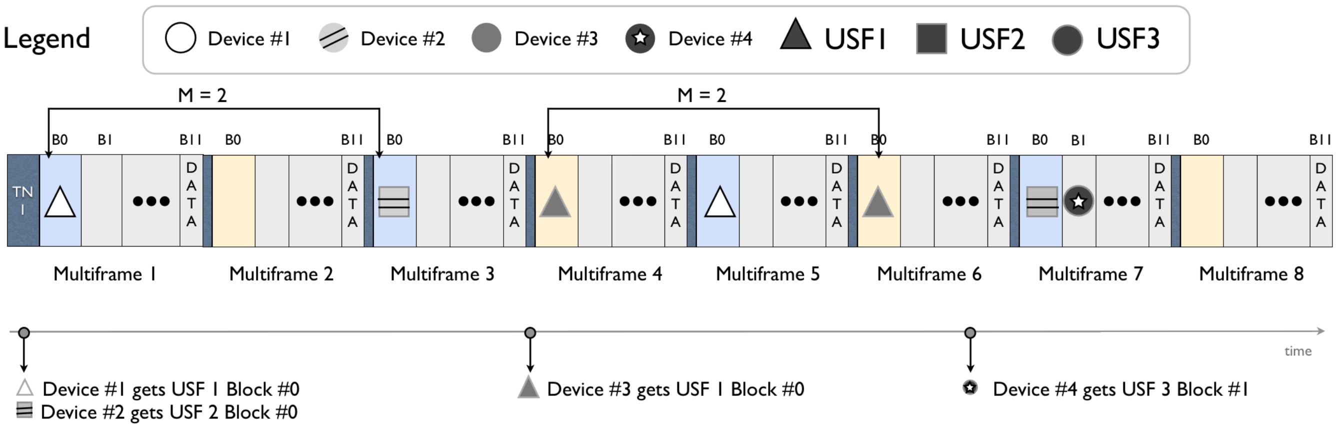

Fig. 4: Example of the proposed expansion of the allocation space.

[11]. The arrival rate is given by $\lambda = 1/RI$, where RI is the report interval. We assume that each device has a packet of constant size RS. Also, each of the devices is allocated a portion of the link-time, as elaborated in Section III. The radio link conditions are assumed to be ideal (i.e., no channel-induced errors are considered).

B. Achievable Capacity in the modified system

The total capacity of the cell (i.e., the number of simultaneous active connections) is determined by the amount of resources needed by each of the devices allocated in the system. The resources are given in terms of blocks needed to transport a report of size RS bytes and the used modulation coding scheme (MCS). Thus, the total capacity is[3]:

$$C = M \cdot \frac{X \text{ Users} \cdot 12 \text{ Blocks} \cdot L \text{ PDCHs}}{\lceil \frac{RS}{MCS} \rceil} \text{ [Devices]}, \quad (1)$$

where L is the number of "data" PDCHs and X the number of USFs available.

In Table I, we show the total cell capacity for different values of M, and the amount of information carried out depending on the coding scheme used and L is the number of PDCHs present in the cell. In addition, the tradeoff between number of devices and the time between transmissions is shown. The larger M is, the more devices are allocated in the system. However, more devices implies larger time between scheduled opportunities T. The duration of this period is given by:

$$T = M \cdot X \cdot 0.24 \text{ [s]} \quad (2)$$

where 0,24 corresponds to the duration of a multiframe expressed in seconds. In addition, the resources given to any device should preserve the reporting interval. Thus, we set a constraint on T:

$$T \leq RI \quad (3)$$

[3]Again, we assume a single frequency.

M Value	Bytes per Transmission MCS1	MCS5	Time Between Scheduled Opportunities (T)	Capacity (devices)
18	22	56	30 s	≈ 10.5 K
35	22	56	≈ 1 min	≈ 20 K
175	22	56	≈ 5 min	≈ 102 K
2100	22	56	≈ 1 h	≈ 1.2 M

TABLE I: Total cell capacity with M configured to example reporting intervals in smart metering when single frequency is used. The assumed values of X and L are 7.

C. Queue Model

The total cell capacity only indicates what is the number of devices provided with a transmission opportunity. However, the actual system capacity depends also on the delay tolerance of the devices, and might significantly differ from the results displayed in Table I. In this subsection we analyze the probability that the report delivery time exceeds the delay tolerance (i.e., loss probability), which can be used to calculate what is the number of devices that can be effectively served.

As elaborated before, we assume that all devices are pre-allocated for service each M multiframes - we assume that the duration of this period is equal to T seconds. Further, we assume that the amount of resources given to a device is limited to a single report, and that the delay tolerance T_d within which the report has to be sent is also equal to reporting interval of the smart meter (RI):

$$T_d = RI \quad (4)$$

any report with delivery time larger than the delay tolerance is dropped and considered as lost.

The above system can be modeled as a multiple vacation queue with limited service and impatient customers (i.e. a queue with reneging). In this system, a customer is only willing to wait up to T_d seconds, after which becomes a lost

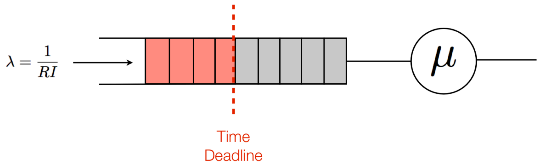

Fig. 5: M/D/1 Queue with impatient customers used to model the system. The service time μ depends on the system parameters M and MCS.

customer. Each time the server visits the queue it serves a single customer if the queue is not empty, and then goes on vacation. If the queue is empty, the server goes to vacation immediately.

The literature on queuing systems with vacations, limited service and reneging is scarce and, to the best of our knowledge, offers no closed form solutions that could be applied to the case in consideration. In order to analytically model the above system, we consider an approximation in which the service time is T seconds. In other words, we approximate a server with vacations with a server with service rate of $\mu = \frac{1}{T}$; as shown later, this approximation will actually yield and upper bound on the loss probability. We note that similar assumptions have been used to model the service rate in GSM in [12].

Without loss of generality we assume that each device sees a dedicated channel, as the service rate does only depend on the parameter M. Thus, the system can be modeled as a queue with a deterministic service time with impatient customers. The number of reports initiated by the device per second is given by a Poisson process $P \sim \text{Poisson}(\lambda)$, where the arrival rate λ is given by the inverse of the reporting interval, i.e., $\lambda = 1/RI$.

The (report) loss probability P_L is the probability that the sum of its waiting time in the queue and the time required for its service exceeds RI, and it can be calculated by the means of the steady-state cumulative distribution function (cdf) of the amount of unprocessed work $U(x)$ as [13]:

$$P_L = 1 - U(\mu\tau) \tag{5}$$

where $\tau = T_d - \frac{1}{\mu}$.

The amount of unprocessed work in the modified queue can be expressed in terms of the amount of unprocessed work in a standard M/D/1 queue U_∞(x) [14]:

$$U(x) = \frac{U_\infty(x)}{1 - \rho + \rho U_\infty(\mu\tau)} \quad \text{for } 0 \leq x \leq \mu\tau \tag{6}$$

where, ρ is the utilization factor equal to $\rho = \frac{\lambda}{\mu}$.

By the PASTA property [15], the amount of unprocessed work $U_\infty(x)$ is identical to the waiting time of a customer when service is first-come, first-served. The waiting time distribution $W_\infty(x)$ in a M/D/1 system can be written as [16]:

$$W_\infty(x) = (1-\rho)\sum_{j=0}^{q}(-1)^j \frac{(\rho x - \rho j)^j}{j!} e^{\rho(x-j)} \tag{7}$$

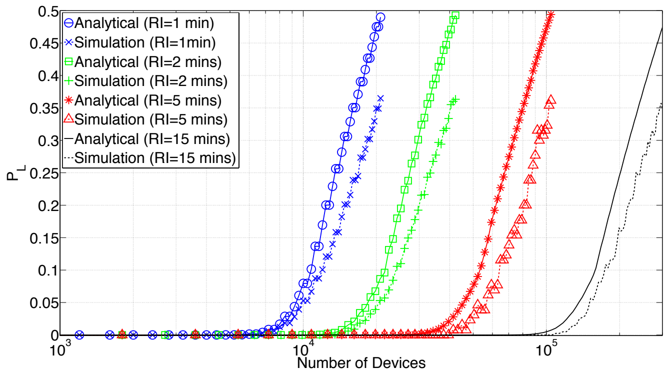

Fig. 6: Comparison between event driven system simulation and the analytical model.

where q is the largest integer less than or equal to x, $q = \lfloor x \rfloor$.

Finally P_L is:

$$P_L = \frac{1}{\rho} \cdot \left[\rho - 1 + \left\{1 + \rho\, e^{(\lambda\tau)} \sum_{j=0}^{q}(-1)^j \frac{(\lambda\tau - \rho j)^j}{j!}\, e^{(-\rho j)}\right\}^{-1}\right] \tag{8}$$

Before we proceed with the performance analysis in Section V, we provide a comparison between a simulation of the proposed system behavior and the presented approximate analysis in Fig. 6. From Fig. 6 it can be observed that the analysis provides an upper bound on P_L.

V. Results

In this section we investigate the inverse relationship between the number of devices and the P_L that can be achieved in a GSM for the example cases of smart meters with report sizes of 100 and 500 bytes and report intervals 1, 2, 5 and 15 minutes. We assume that the delay tolerance is equal to RI and that the devices use modulation coding scheme MCS-5 (56 bytes per block).

Fig. 7 presents the results when report size is 100 bytes. It is noticeable that, for the most demanding case when RI=1min, a single cell could provide service for up to $5{\cdot}10^3$ simultaneous connections with a reliability of 99.99%. This number rises to outstanding value of $5 \cdot 10^4$ simultaneous connections that are served with 99.99%, if the reporting interval is set to 15 min.

Finally, the scenario when the report size is 500 bytes is illustrated in Fig. 8. The effects of carrying larger messages are obvious, as compared to Fig. 7; now the base station has to provide more blocks per device, which decreases the capacity. Nevertheless, the system is able to provide service to 10^4 active devices with at least one transmission opportunity each 15 min.

Finally, we note that the above results represent lower bounds on the number of devices that could be supported in a cell; for the given P_L the actual number could be even higher (see Fig. 6).

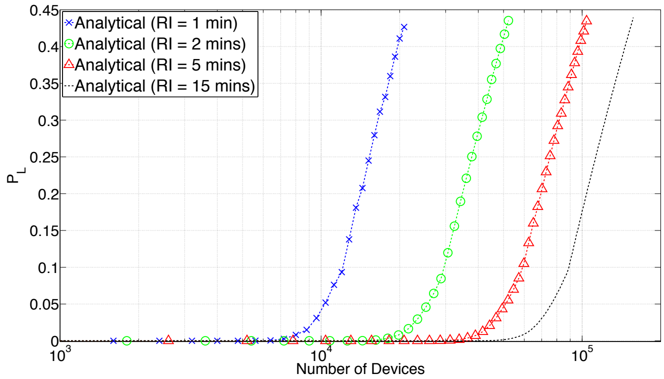

Fig. 7: Loss probability P_L as function of report interval RI, report size 100 bytes.

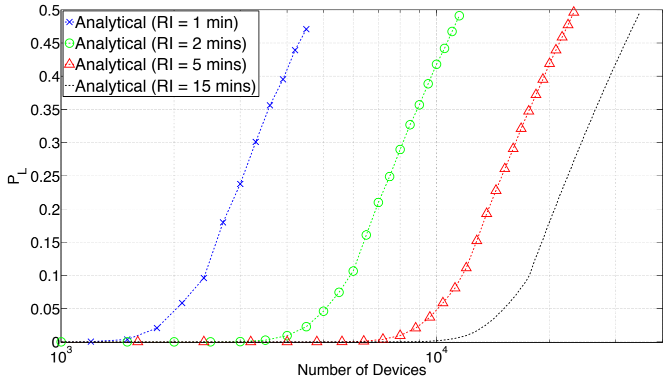

Fig. 8: Loss probability P_L as function of report interval RI, report size 500 bytes.

VI. Conclusions

In this work, we presented a method to reengineer the way devices are allocated resources in GSM, providing for much finer granularity and thus extensively boosting the capacity of the system. Particularly, we have shown that the proposed method for expanding the allocation space allows for support of number of devices that is several orders of magnitude higher than it is achievable in the standard GSM use. The suggested modifications incur only modest changes at the MAC layer, while the physical layer (i.e., air interface) remains unaltered. Moreover, the proposed solution can coexist with the current standard - a subset of the available PDCHs in the cell can be operated in the proposed fashion, while the rest can be operated in the standard way. Further, we demonstrated the performance of the proposed method for the example case of smart metering applications, where the delivery of meter reports is subject to deadlines, by deriving the analytical bounds on the report loss probability and calculating the corresponding call capacities. The obtained results show the considerable potential of GSM to serve as a carrier for smart metering applications.

Acknowledgment

The research presented in this paper was supported by the Danish Council for Independent Research (Det Frie Forskningsråd) within the Sapere Aude Research Leader program, Grant No. 11-105159 Dependable Wireless bits for Machine-to-Machine (M2M) Communications.

References

[1] Q. D. Vo, J. P. Choi, H. M. Chang, and W. C. Lee, "Green perspective cognitive radio-based m2m communications for smart meters," in *Information and Communication Technology Convergence (ICTC), 2010 International Conference on*. IEEE, 2010, pp. 382–383.
[2] L. X. D. Niyato and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.
[3] *Sierra Wireless Product Webpage.*, Accessed on October 2012, http://www.sierrawireless.com/en/Solutions/customer_stories/EDMI.aspx.
[4] *Quebec Press Release Smart Metering.*, Accessed on December 2012, https://www.rogers.com/web/link/showNewsDetail?fromWhere=linkInRSSXml&rssBusiUnit=W&NewsID=1906177072.
[5] *Endesa Press Release Smart Metering.*, Accessed on January 2012, http://www.endesasmartgrid.com/index.php/en/smart-homes/smart-metering.
[6] 3GPP, "Service Requirements for Machine-Type Communications (Stage 1)," 3rd Generation Partnership Project (3GPP), TS 22.368, Jun. 2010. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23368.htm
[7] 3GPP, "Bottleneck Capacity Comparison for MTC," 3rd Generation Partnership Project (3GPP), TSG GERAN #46 GP-100893, 2010.
[8] 3GPP, "RACH intensity of Time Controlled Devices," 3rd Generation Partnership Project (3GPP), TSG RAN WG2 R2-102340, 2010.
[9] 3GPP, "RACH Capacity Evaluation for MTC," 3rd Generation Partnership Project (3GPP), TSG GERAN #46 GP-100892, 2010.
[10] 3GPP, "USF Capacity Evaluation for MTC," 3rd Generation Partnership Project (3GPP), TSG GERAN #46 GP-100894, 2010.
[11] IEEE 802.16p, "IEEE 802.16p Machine to Machine (M2M) Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group (802.16p), EMD 11/0014, May 2011.
[12] H. Dahmouni, B. Morin, and S. Vaton, "Performance modeling of gsm/gprs cells with different radio resource allocation strategies," in *Wireless Communications and Networking Conference, 2005 IEEE*, vol. 3, march 2005, pp. 1317 – 1322 Vol. 3.
[13] A. De Kok and H. Tijms, "A queueing system with impatient customers," *Journal of applied probability*, pp. 688–696, 1985.
[14] H. Tijms, *A first course in stochastic models*. Wiley, 2003.
[15] W. W. Ronald, "Poisson arrivals see time averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, 1982.
[16] V. Iversen and L. Staalhagen, "Waiting time distribution in M/D/1 queueing systems," *Electronics Letters*, vol. 35, no. 25, pp. 2184–2185, 1999.