<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *VTC 2017 Spring*.

Citation for the original published paper:

Azari, A., Istiak Hossain, M., Markendahl, J. (2017)
RACH Dimensioning for Reliable MTC over Cellular Networks.
In: *VTC 2017 Spring* Sydney

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-200680>

RACH Dimensioning for Reliable MTC over Cellular Networks

Amin Azari, Mohammad Istiak Hossain, and Jan I Markendahl

KTH Royal Institute of Technology

Email: {aazari, hossain7, janmar}@kth.se

Abstract—Successful implementation of massive machine-type communications (MTC) service over cellular networks is an important enabler of Internet of things. As the existing cellular infrastructure has been designed and optimized for moderate traffic, the introduction of MTC in cellular networks results in several key challenges, among them, temporal starvation of initial access resources is of paramount importance. Here, we propose a new traffic model for cellular networks serving machine-type subscribers. This model is subsequently used to derive closed-form expressions for the access rate performance of the system. This closed-form expression are subsequently used for RACH resources planning in the time and frequency domains. Given a target access rate requirement, numerical analysis indicates how and when the required RACH resources in the code/frequency domain, i.e. number of preambles, can be traded with the occurrence frequency of RACH slots in the time domain.

Index Terms—M2M, MTC, IoT, Massive access, 5G, Cellular networks.

I. INTRODUCTION

Current market research anticipated an exponential growth of internet of things (IoT) service set in coming years. IoT is referred to massive autonomous connected devices' services that lead to novel communications business opportunities. Formerly, IoT connectivity was foretold as short range, multi hop and low power systems. However, presently there is a strong drive for cellular IoT connectivity. Naturally, this new service set opens up a whole new revenue stream for mobile network operators (MNOs). This brings the machine to machine (M2M) communications as a service in order to facilitate the connectivity of IoT devices. M2M also frequently stated as machine-type communications (MTC) or cellular IoT is the communications between autonomous machines without human intervention to control and monitor applications. The term MTC is usually used when the device is connected through a cellular network. This MTC service is expected to play a key role in successful deployment of IoT due to existing ubiquitous coverage, secure connection and roaming capability of cellular networks like LTE, LTE-A and future 5G networks.

The service applications and use cases of future cellular Networks services are wide-ranging from conventional service set like mobile broadband (MBB) and voice services (i.e. LTE-VOIP) to 4k video streaming, augmented reality and massive IoT-type device connections [1]. This multi-service traffic will be generated by large numbers of asynchronous¹ devices belong to different service sets. Also, device density per area will increase in the order of couple of hundred

thousand² per km^2 mainly due to the rise of cellular IoT services. This in turn means the amount of asynchronous multi-service generated traffic per base station (BS) will be very high. Then, the well-known traffic and network access patterns that go all through the network will change. In such condition, the more various types of services will introduce to the cellular network, the more bursty access pattern will spawn [2]. Additionally, event driven IoT services like smart grid, disaster management, earthquake detection, flood detection services may cause access crunch by triggering many devices in a geological location to send incident report to the server. In existing cellular networks, a user device may allow transmitting and receiving data if it is time-synchronized with the BS and has received resource grant from the BS. The main role of random access (RA) procedure in cellular networks is to request this initial access resource grant and establish uplink synchronization between device and BS [3].

The next generation cellular network (5G) is aimed to support large numbers of service sets than existing cellular networks like LTE and LTE-A. One of the main envisioned features of 5G is to handle 1000 times higher device density per area than existing systems [4]. Presently, cellular networks physical random access channel (PRACH) is dimensioned and utilized to support moderate access loads [5]. Massive MTC devices that transmit small payloads periodically may not create resource scarcity. Though, they may interrupt other services because of the plurality of present devices in a certain area, and the probability of triggering a massive concurrent access event. The current random access protocol has two main limitations for bursty-type access/device arrivals. First, number of random access opportunities (RAOs) per cell is limited. Second, number of physical downlink control channel (PDCCH) resources, which are used to send Ack/Nack/access grant messages, is limited. These two bottlenecks limit the number of concurrent devices that can be served at a time.

3GPP and ETSI standardization of LTE has considered analytical preamble collision probability models for PRACH dimensioning [3gpp R1-061369]. In literature, scientific papers like [3]- [7] have investigated preamble collision probability to estimate device success rate and network dimensioning. In [8] an analytical model is presented to estimate the performance of the RACH procedure in case of massive concurrent access request arrival. However, we noticed that none of these papers have considered retransmissions impact on access load in their model. The authors in [9] has proposed a model and

¹asynchronous from service access requirement perspective because in multi-service setting, services dont do cooperative network access.

²in ultradense scenario

Fig. 1. HTC and MTC Traffic patterns [10], [11].

considered the retransmission impact, but they failed to derive a closed-form expression for the access rate performance. Furthermore, the traffic model used in [9] and most existing literature is Poisson arrival. We will argue in section III that this model is not a good match for MTC/Humane-type communications (HTC) traffic coexistence, and the derived results by using this model will be either upper- or lower-bounds on the system performance, based on how one tunes the Poisson process' parameter. However, [2] provides a real trace of generated traffic from MTC applications. This trace shows periodic bursts of MTC traffic over HTC traffic. In this paper, we introduce an analytic traffic model for cellular networks serving multi-service traffic, which captures the characteristics of the presented trace in [2]. This model is subsequently used to derive closed-form expressions for the access rate performance of the system. These expressions are very useful in network dimensioning. We may use these expressions for RACH resource dimensioning in time and frequency domains. The main contributions of this paper include:

- Develop a tractable framework to model the RACH performance in case of massive arrival considering re-transmissions and access class barring. Propose a new traffic model capturing the characteristic of massive MTC traffic, including regular reporting and event-driven, over cellular network.
- Derive closed-form accessibility expressions as a function of traffic parameters e.g. length and intensity of the massive arrival mode, resource planning parameters e.g. number and frequency of occurrence of RACH resources, and RACH parameters e.g. backoff duration. Investigate the impact of planned RACH resources on the reliability of network, and figure out the tradeoffs.

The rest of this paper is organized as follows. System model is presented in the next section. The proposed traffic model and problem formulation are investigated in section III. Access rate analysis is presented in section IV. Numerical and simulation results are presented in section V and VI. The concluding remarks are given in section VII.

II. SYSTEM MODEL

Consider a single cell with one base station at the center, and a massive number of user equipments (UEs) which are uniformly distributed in the cell. Random access channel (RACH)

of LTE is the typical way for inactive devices to establish connection with the BS. Regarding the contention-based nature of RACH, collision in random access is inevitable, and hence, access class barring has been standardized in cellular networks. Based on the ACB, each device which has data to transmit decides to start contention at each time slot with probability q or postpone the contention with probability $1 - q$. If a device makes K successive unsuccessful attempts, it will be blocked. As highlighted by the Metis project [12] and state-of-the-art [13], reliability in communications is of paramount importance for future cellular networks. For example, in test-case 5 of [12], which deals with teleprotection for smart grid networks, 99.99 reliability has been envisioned. Motivated by the facts that the number of random access resources per cell per time is limited and the disconnected devices must go through the RACH procedure to transmit data to the BS, RACH resource planning is of paramount importance in order to guarantee a required level of access rate to the network, while preserving the expensive radio resources. To achieve this level of reliability, in the following sections we investigate the ways in which RACH planning in time and frequency can be performed to ensure a required level of accessibility to the cellular network.

III. TRAFFIC AND ACCESS-RATE MODELING WITH HTC/MTC COEXISTENCE

Access rate, denoted by R , demonstrates the percentage of devices that can successfully pass the RACH procedure within a limited number of retransmission opportunities. Uplink RACH resources, red-colored blocks in Fig. 2(a), are reserved to occur in regular intervals of T_{RA} seconds. Each RACH slot in frequency domain consists of M_p orthogonal preambles, and M_f parallel RACH slots can be planned in frequency domain per time slot. Then, $M = M_f M_p$ preambles have been reserved in each T_{RA} seconds for random access. Other uplink resources are reserved for physical uplink shared/control channels (PUSCCH and PUSCH).

A. The Proposed Traffic Model

When there is only HTC traffic, the traditional Poisson user arrival model can be used for performance modeling, as depicted in Fig. 2(b). While in case of MTC/HTC coexistence, specially when event-driven MTC traffic is taken into account, the achieved access rate from Poisson model cannot be used anymore. A switched-Poisson process (SPP) model matches well with the new arrivals of HTC/HTC traffic to the system, i.e. when there is no retransmissions, as depicted in Fig. 2(c). SPP introduces high- and low-arrival rate modes for the traffic. The lengths of high- and low-arrival-rate (HAR and LAR) modes can be modeled as $L_h = n_h \tau$ and $L_l = n_l \tau$, in which n_h and n_l are geometrically distributed with parameter μ_h and μ_l respectively. While the SPP model can model the arrival rate of users, retransmission after failed transmission poses challenges in modeling the access rate performance as it adds a transient state to the system in which a complex set of new-arriving/re-attempting nodes try to access the network. Then,

Fig. 2. RACH resource planning and traffic modeling

in order to model the traffic a modified SPP model is proposed in Fig. 2(d) which consists of the HAR, LAR, and transient states. The state transitions of the proposed model are depicted in Fig. 3. As the worst case access rate is achieved in the case of massive access occurrence, in the following we focus on the following transition: (LAR → HAR → transient states).

IV. ANALYTICAL PERFORMANCE ANALYSIS

Denote granularity of radio resources in time domain as τ , the time interval between two successive random access opportunities (RAO) as T_{RA} , the ratio between T_{RA} and τ as $\mathcal{X} \geq 1$, number of RACH frequency slots per RAO as M_f , number of orthogonal preambles in each RACH slot as M_p , and maximum number of allowed (re-)transmissions as K . Let us assume that the high-arrival-rate mode starts at the reference time, i.e. t_0 . The expected number of attempting devices at t_1 is derived as

$$\begin{aligned} \mathcal{N}_1 &= \text{arrived in } [t_0, t_1] + \text{failed at } t_0 \\ &\stackrel{(a)}{\approx} \text{arrived in } [t_0, t_1], \end{aligned} \quad (1)$$

where (a) is due to the fact that massive access occurs at t_0 . Also, $\mu_h \gg \mu_l$ is assumed due to the characteristics of the IoT traffic [10]. Then, $\mathcal{N}_1$ can be approximated as:

$$\begin{aligned} \mathcal{N}_1 &\approx \lambda_h T + \sum_{i=0}^{\mathcal{X}-1} [\lambda_h i\tau + \lambda_l (\mathcal{X} - i - 1)\tau] (1 - \mu_h)^i \mu_h^{I(\mathcal{X}-i-1)}, \\ &\approx \lambda_h \min\{T_{RA}, \frac{\tau}{\mu_h}\} + \lambda_l \max\{T_{RA} - \frac{\tau}{\mu_h}, 0\}, \end{aligned}$$

where $I(z) = 1$ for $z > 0$ and 0 otherwise. Regarding ACB,

$$N_1 = \mathcal{N}_1 (1 - (1 - q)^X)$$

devices out of $\mathcal{N}_1$ will decide to contend at t_1 , where $1 - (1 - q)^X$ is the probability that a device decides to

Fig. 3. State transitions of the proposed model for the coexistence traffic. The below states are transient states.

content for channel access after X time slots, and q shows the probability of decision for contention in each time slot. The average probability of collision in random access for x devices and y preambles is derived as:

$$P_c(x, y) = 1 - \left[\frac{y - 1}{y} \right]^{x-1}.$$

Then, $N_1 P_c(N_1, M)$ devices will be unsuccessful at t_1 and will retry at t_2 , where $M = M_p M_f$. Also, some new devices will try for random access at t_2 . The number of new arrivals depends on the length of HAR mode, i.e. $l_h = n_h \tau$. Regarding the geometric distribution of n_h , the average number of new arrivals can be modeled as

$$\begin{aligned} \mathcal{N}_2^{new} &= \lambda_l \mathcal{X} \tau [1 - \sum_{i=1}^{\mathcal{X}} (1 - \mu_h)^{\mathcal{X}+i-1} \mu_h^{I(\mathcal{X}-i)}] + \\ &\sum_{i=1}^{\mathcal{X}} [\lambda_h i\tau + \lambda_l (\mathcal{X} - i)\tau] (1 - \mu_h)^{\mathcal{X}+i-1} \mu_h^{I(\mathcal{X}-i)}. \end{aligned} \quad (2)$$

Then, at time t_2 ,

$$\mathcal{N}_2 = \mathcal{N}_2^{new} + \mathcal{N}_1 [(1 - (1 - q)^X) P_c(\mathcal{N}_1, M) + (1 - q)^X]$$

devices will have data to transmit, out of which $N_2 = \mathcal{N}_2 (1 - (1 - q)^X)$ will decide to contend for RACH transmission. Following the same trend, at time t_k , where $k > 1$,

$$\mathcal{N}_k = \mathcal{N}_k^{new} + \mathcal{N}_{k-1} [(1 - (1 - q)^X) P_c(\mathcal{N}_1, M) + (1 - q)^X]$$

devices will have data to transmit, where

$$\begin{aligned} \mathcal{N}_k^{new} &= \lambda_l \mathcal{X} \tau [1 - \sum_{i=1}^{\mathcal{X}} (1 - \mu_h)^{(k-1)\mathcal{X}+i-1} \mu_h^{I(\mathcal{X}-i)}] + \\ &\sum_{i=1}^{\mathcal{X}} [\lambda_h i\tau + \lambda_l (\mathcal{X} - i)\tau] (1 - \mu_h)^{(k-1)\mathcal{X}+i-1} \mu_h^{I(\mathcal{X}-i)}. \end{aligned}$$

Then, $N_k = \mathcal{N}_k (1 - (1 - q)^X)$ devices will decide to contend at t_k . Given K , the maximum number of attempts, for delay-critical traffic q must be selected such that on average K attempts fall inside the tolerated delay in RACH access, i.e. T_d , as follows:

$$\lceil T_d / X\tau \rceil / \lceil (1/q) / X \rceil \geq K,$$

where $\lceil x \rceil$ shows the ceiling of x . One sees that access rate for a given device is a function of probabilities of collision in the selected set of RAOs in which the device contend for channel access. The expected access rate for devices that start contention at t_j is calculated as:

$$\begin{aligned}
R_j &= 1 - \prod_{l=0}^{K-1} P_c \left(N_{j+l} \left\lceil \frac{1/q}{\mathcal{X}} \right\rceil, M \right) \\
&= 1 - \prod_{l=0}^{K-1} \left(1 - \left[\frac{M-1}{M} \right]^{N_{j+l} \left\lceil \frac{1/q}{\mathcal{X}} \right\rceil} \right), \forall j \geq 1. \quad (3)
\end{aligned}$$

For traffic models in which the expected duration of HAR mode is less than or equal to one T_{RA} , i.e. $\mu_h > \mathcal{X}^{-1}$, the worst case access rate is given by R_1 . For example, given $M, T_{RA}, \lambda_h, \lambda_l, \mathcal{X} = q = 1$, and $K = 2$, we have:

$$\begin{aligned}
R_1 &= 1 - P_c(N_1, M) P_c(N_2, M) \quad (4) \\
&\approx 1 - \left(1 - \left[\frac{M-1}{M} \right]^{\lambda_h T_{RA}-1} \right) \left(1 - \left[\frac{M-1}{M} \right]^{(\lambda_h T_{RA}) \dots} \right. \\
&\quad \left. \dots \times \left(1 - \left[\frac{M-1}{M} \right]^{\lambda_h T_{RA}-1} \right) + \mu_h \lambda_l T_{RA} + (1-\mu_h) \lambda_h T_{RA} - 1 \right) \\
&\approx 1 - (1 - \mathcal{M}^{\lambda_h T_{RA}}) \times \\
&\quad \left(1 - \mathcal{M}^{(1-\mu_h) \lambda_h T_{RA} + \mu_h \lambda_l T_{RA} + \lambda_h T_{RA} (1 - \mathcal{M}^{\lambda_h T_{RA}})} \right), \quad (5)
\end{aligned}$$

in which, $\mathcal{M} = \frac{M-1}{M}$. Given $T_{RA}, \lambda_h, \lambda_m$, and the required access rate, one sees that the minimum required number of preambles for satisfying the access rate, i.e. M , can be derived from (5). The expressions in (3)-(5) represent the impact of RACH resource allocation on the access rate. It is evident that R_1 is an increasing function of M and a decreasing function of T_{RA} , i.e. the access rate increases when the amount of RACH resources in each random access opportunity (RAO) increases, and decreases when the time interval between two successive RAOs increases. The expression in (3) can also be used for RACH planning, i.e. the required set of resources for achieving a target access rate, e.g. 99.99%, can be derived from (3). For traffic models in which the HAR mode spans over more than one T_{RA} , i.e. $\mu_h > \mathcal{X}^{-1}$, the worst case access rate is

$$\min_{j \in \left\{ 1, \dots, \left\lceil \frac{1}{\mathcal{X} \mu_h} \right\rceil \right\}} R_j.$$

V. NUMERICAL ANALYSIS: RACH RESOURCE PLANNING FOR CELLULAR-MTC

Here, we investigate RACH resource planning using MATLAB. Fig. 4 depicts access rate as a function of number of preambles and the time interval between two successive RAOs. The system and traffic parameters can be found in Table I. It is evident that a target access rate can be achieved by adequate resource allocation in time and frequency domains. For example, one sees that 0.9999 access rate is achieved both with $(M = 161, T_{RA} = 1 \text{ msec})$ and $(M = 261, T_{RA} = 2 \text{ msec})$. Then, one may decrease the allocated bandwidth to RACH by increasing the occurrence frequency of RAOs in the time domain. The detailed access rate analysis is given in Fig. 5, in which the yellow-colored regions specify the (M, T_{RA}) values that satisfy the access rate requirement. It is interesting to see that, given the access rate, the boundary values of M and T_{RA} that satisfy the access rate construct an exponential curve. Doing a linear approximation of this curve, one sees that the slope of this curve is inversely proportional to the access rate,

Fig. 4. Access rate as a function of number of preambles in each RA (M) and the time interval between two successive RAOs (T_{RA}). Parameters: $q = 1$, $\tau = 1 \text{ msec}$, $\mu_h = 0.05$, $\lambda_m = 100$, and $\lambda_h = 2000$.

Fig. 5. Access rate as a function of number of preambles in each RA (M) and the time interval between two successive RAOs (T_{RA}). The threshold access rates are 0.9999, 0.99, and 0.95. The yellow regions show the configurations in which the target access rate has been satisfied.

i.e. by increase in the access rate requirement the substitution rate of M with T_{RA} increases. This in turn indicates that for networks with a high access rate requirement, one may not achieve the target access rate only by decreasing the interval between two successive occurrences of RAOs. In this case, increasing both the number of preambles per RAO and the occurrence frequency of RAOs are required.

VI. SIMULATION RESULTS

We validate our analytical estimation with our preliminary simulation results. We use NS3 LENA model [14] for RA performance assessment in LTE-FDD mode for a single cell scenario, where PRACH period is set to 10ms and PRACH configuration Index is 4. In uplink direction, a coexistence of MTC and HTC traffic has been implemented. The UEs are randomly distributed according to a spatial Poisson point process in the cell. Access request from UEs arrive based on the modified SPP distribution. It is also assumed that the eNB will always be able to detect preamble collision and does not respond to that request. We evaluate the performance by using the simulation parameters presented in Table I. We ran the test 25 times for each maximum-allowed-number-of-retransmission (MANR) and resource slots iteration. The simulations results along with the corresponding analytical results have been depicted in Fig. 4. From this figure, one can see that the simulation results follow the same trend as the numerical results, specially when MANR is larger than 7, where the numerical and simulations results are exactly the same. Also, one sees that there is a short gap between the analytical and simulations results when MANR is less than equal 7. However, as the curve resulting from the simulation results in this region is twisted to the curve resulting from the numerical results, we expect the two curves to match even better by increasing the number of iterations.

VII. CONCLUSION

Challenges in RACH planning with introduction of massive MTC services to cellular networks have been investigated. First, a novel traffic model has been proposed for cellular networks serving multi-service traffic, which is able to capture the main characteristics of massive MTC traffic, specially the sporadic nature of massive concurrent access arrival occurrence. Based on this model, closed-form expressions have been derived to analyze the access rate performance of the system for delay critical services, in which the number of retransmission attempts is limited. Numerical and simulations analyses represent ways in which the derived expressions can be used for proper RACH planning in time and frequency domains.

TABLE I
SIMULATION PARAMETERS

Parameter	Values	Unit
Devices per cell	3000	int.
No. Available preambles	54	int.
PRACH configuration Index	4	int.
Backoff Indicator	0	ms
PRACH period	10	ms
Max Preamble retransmissions	3,5,7,9,11	int.
RAR window size	3	ms
Connection request timeout	60	ms
Cell radius	2	km
MTC device Ratio	70	%
Arrival rate in HAR mode	20000	per sec.
Arrival rate in LAR mode	100	per sec.
Avg. HAR Duration	20	ms

Fig. 6. Comparison between analytical estimation and simulation result of resource requirements to achieve 99.99% access rate

ACKNOWLEDGMENT

This work has been partially funded by Wireless@kth and H2020 project METIS-II (H2020-ICT-2014-2, topic ICT-14-2014).

REFERENCES

- [1] METIS-I project, "Refined scenarios and requirements, consolidated use cases, and qualitative techno-economic feasibility assessment," December 2014.
- [2] M. Z. Shafiq *et al.*, "A first look at cellular machine-to-machine traffic: large scale measurement and characterization," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 1, pp. 65–76, 2012.
- [3] C. Ubeda *et al.*, "LTE FDD physical random access channel dimensioning and planning," in *Vehicular Technology Conference (VTC Fall)*, 2012 IEEE. IEEE, 2012, pp. 1–5.
- [4] H2020 METIS-II project, "5G RAN architecture and functional design," February 2016.
- [5] METIS-I project, "Scenarios, requirements and kpis for 5g mobile and wireless system," January 2015.
- [6] R. G. Cheng *et al.*, "RACH collision probability for machine-type communications," in *VTC Spring, 2012 IEEE 75th*. IEEE, 2012, pp. 1–5.
- [7] C. Karupongsiri *et al.*, "Random access issues for smart grid communication in lte networks," in *ICSPCS, 2014 8th International Conference on*. IEEE, 2014, pp. 1–5.
- [8] O. Arouk and A. Ksentini, "General model for rach procedure performance analysis," *IEEE Communications Letters*, vol. 20, no. 2, pp. 372–375, 2016.
- [9] J. J. Nielsen *et al.*, "A tractable model of the lte access reservation procedure for machine-type communications," in *IEEE GLOBECOM 2015*. IEEE, 2015, pp. 1–6.
- [10] V. B. Mišić, J. Mišić *et al.*, "M2M traffic and models," in *Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications*. CRC Press, 2014, pp. 57–85.
- [11] M. Laner *et al.*, "Traffic models for machine type communications," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems*. VDE, 2013, pp. 1–5.
- [12] C. S. Petar Popovski *et al.*, "Requirement analysis and design approaches for 5G air interface," Tech. Rep., 08 2013.
- [13] J. J. Nielsen *et al.*, "A tractable model of the LTE access reservation procedure for machine-type communications," in *IEEE Global Communications Conference*, 2015, pp. 1–6.
- [14] N. Baldo *et al.*, "An open source product-oriented LTE network simulator based on ns-3," in *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*. ACM, pp. 293–298.