Aalborg Universitet

A Tractable Model of the LTE Access Reservation Procedure for Machine-Type Communications

Jimmy J. Nielsen, Dong Min Kim, Germán C. Madueño, Nuno K. Pratas, Petar Popovski
APNET section, Department of Electronic Systems, Aalborg University, Denmark
{jjn,dmk,gco,nup,petarp}@es.aau.dk

Abstract—A canonical scenario in Machine-Type Communications (MTC) is the one featuring a large number of devices, each of them with sporadic traffic. Hence, the number of served devices in a single LTE cell is not determined by the available aggregate rate, but rather by the limitations of the LTE access reservation protocol. Specifically, the limited number of contention preambles and the limited amount of uplink grants per random access response are crucial to consider when dimensioning LTE networks for MTC. We propose a low-complexity model of LTE's access reservation protocol that encompasses these two limitations and allows us to evaluate the outage probability at click-speed. The model is based chiefly on closed-form expressions, except for the part with the feedback impact of retransmissions, which is determined by solving a fixed point equation. Our model overcomes the incompleteness of the existing models that are focusing solely on the preamble collisions. A comparison with the simulated LTE access reservation procedure that follows the 3GPP specifications, confirms that our model provides an accurate estimation of the system outage event and the number of supported MTC devices.

I. INTRODUCTION

Machine-Type Communication (MTC) is commonly characterized by a large number of cellular devices that are active sporadically, where a large number of devices may activate in a correlated way due to a sensed physical phenomenon (e.g., a power outage in the smart grid). In more traditional human-centric traffic where the associated payloads are relatively large, a small number of active devices can cause the network to become in outage mainly due to the lack of available resources for data transmission. In contrast, the associated payloads are relatively small in MTC such that the division of the aggregate available data rate with the small data rate required by each Machine-Type Device (MTD) leads to the conclusion that the system can support a vast number of MTDs. Recent studies have shown that such a conclusion is misleading: the network still becomes in outage, not being able to provide access to the MTDs, despite plenty of available resources to support a massive number of MTDs. Here the culprit in the limited number of supported devices, is not the available resources as in human-centric traffic, but instead the bottlenecks in the access reservation protocol [1]. Specifically in LTE, the Access Reservation protocol that is outlined in Fig. 1 has two limitations that unveil with MTC. The first is in *MSG 1*, where only a limited number of preambles can be used to signal a sporadic request for uplink resources to the eNodeB, in the RACH phase. The second is in *MSG 2*, where a bottleneck may be caused by the limited amount of feedback
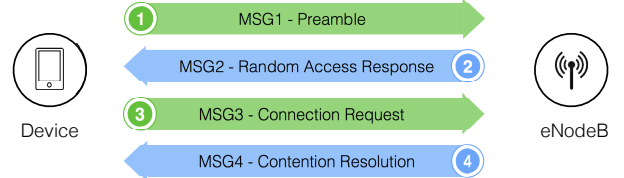

Fig. 1. Message exchange between a device and the eNodeB during the LTE random access procedure.

resources in the access granting (AG) phase.

In the literature, analytical models of the preamble collision probability have already been considered in standardization documents [2]–[4] and scientific papers [5], [6]. In [7] the preamble collision probability is used to estimate the success probability of transmission attempts. However, we have found that existing models are incomplete and inaccurate and in this paper we introduce a superior model that closely matches the system outage breaking point of the detailed simulation.

The second limitation in the AG phase has been considered separate from collisions in [8] for bursty arrivals following the Beta distribution, which is a valuable result for situations where many alarm messages are sent simultaneously. In [6] the authors present an approach to cell planning and adaptation of PRACH resources that only takes into account the preamble collisions. As we show in this paper, the AG phase is a limiting factor before the amount of preamble collisions becomes an issue, since the impact of occasional collisions is effectively diminished with retransmissions. In [9] the authors present an analysis accounting for preamble collision and the AG phase, which however does not consider retransmissions.

In this paper we propose an analytical model of the transmission failure probability in an LTE cell for sporadic uplink transmissions carried over the LTE random access channel. The proposed model captures the features of the existing access reservation protocol in LTE, meaning that we are not proposing a new access protocol rather introducing a tool for analysis of the existing LTE access reservation protocol. The purpose of the proposed model is to be able to estimate the capacity in terms of the number of terminals or RACH arrival density that can be supported by LTE in a given configuration while accounting for retransmissions as well as modeling the bottlenecks that appear in the contention phase and the AG phase. This is a major contribution of the paper, as the existing models do not capture these bottlenecks. Three other
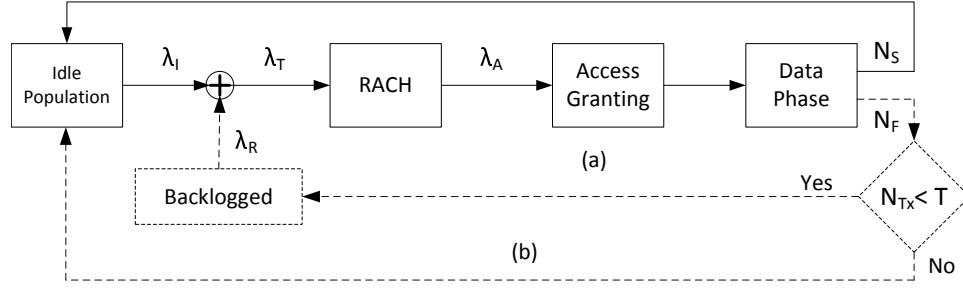
Fig. 2. Flow diagram of LTE access reservation protocol: (a) one-shot transmission model; (b) m-Retransmissions model (dashed lines).

contributions are: 1) an iterative procedure to determine the impact of retransmissions using a Markov chain model of the retransmission and backoff procedure; 2) analytical derivations of the metrics based on a Markov chain, thereby achieving an analytical model that can be evaluated at click-speed. 3) analysis of the protocol breaking point using ever increasing access loads to the network.

Initially we present the system model and assumptions in Section II, whereafter in Section III we present the proposed analytical model of the access reservation bottlenecks in LTE. The proposed model is compared numerically to simulation results and other models from the literature in Section IV, and finally the conclusions are given in Section V.

II. SYSTEM MODEL

We focus our analysis on a single LTE cell, with N MTDs, also called machine-type User Equipment (UE). We assume that the MTC applications associated with these MTDs, generate new uplink transmissions with an aggregate rate λ_I, as depicted in Fig. 2. That is, $\lambda_I = \sum_{i=1}^{K} \lambda_i^{\mathrm{app}}$, where λ_i^{app} is the transmission generation rate of the ith out of K MTC applications running on the UEs. We assume this aggregate rate follows a Poisson distribution with rate λ_I. For each new data transmission, up to m retransmissions are allowed, resulting in a maximum of $m+1$ allowed transmissions. When these transmissions fail and retransmissions occur, then an additional stress is put on the access reservation protocol, since the rate of retransmissions λ_R adds to the total rate λ_T.

As shown on Fig. 2, we split the access reservation model into two parts: (i) the one-shot transmission part in Fig. 2(a) that models the bottlenecks at each stage of the access reservation protocol; (ii) the m-retransmission part in Fig. 2(b), where a finite number of retransmissions and backoffs are modeled. We focus our analysis on MTC, for which the traffic is characterized by having a very small payload. Therefore, in the one-shot transmission, depicted in Fig. 2(a), we assume that the RACH and access granting phases are the system bottlenecks. In other words, we assume that the network has enough data resources to deliver the serviced MTC traffic.

A. LTE Access Reservation Protocol

The uplink resources in LTE for frequency division duplexing (FDD) are divided into time and frequency units denoted resource blocks (RBs) [10]. The time is divided in

frames, where every frame has ten subframes, each subframe of duration $t_s = 1$ ms. The system bandwidth determines the number of RBs per subframe that ranges between 6 RBs and 100 RBs. The number of subframes between two consecutive random access opportunities (RAOs) denoted δ_{RAO} varies between 1 and 20. Every RAO occupies 6 RBs and up to 1 RAO per subframe is allowed.

The LTE random access follows the access reservation principle meaning that devices must contend for uplink transmission resources in a slotted ALOHA fashion within a RAO [11], [12]. As shown in Fig. 1, the access procedure consists of the exchange of four different messages between a UE and the eNodeB. The first message (MSG 1) is a randomly selected preamble sent in the first coming RAO. In Fig. 2(a) the intensity of UE requests leading to preamble activations is represented by λ_T. LTE has 64 orthogonal preambles, where only $d = 54$ are typically available for contention among devices, since the rest are reserved for timing alignment. Commonly, the eNodeB can only detect which preambles have been activated but not if multiple activations (collisions) have occured. This assumption holds in small cells [13, Sec. 17.5.2.3], and refers to the worst-case scenario where the detected preamble does not reveal anything about how many users are simultaneously sending that preamble[1]. In other words, the preamble collision is not detected at MSG 1.

Thereafter, in MSG 2, the eNodeB returns a random access response (RAR) to all detected preambles. The intensity of activated preambles is in Fig. 2(a) represented by λ_A, where $\lambda_A \leq \lambda_T$ since in a preamble collision only 1 preamble is activated. The contending devices listen to the downlink channel, expecting MSG 2 within t_{RAR}. It should be noted that typically a maximum of 3 RAR messages per subframe can be sent by the eNodeB [8]. If no MSG 2 is received and the maximum of T MSG 1 transmissions has not been reached, the device backs off and restarts the random access procedure after a randomly selected backoff time within the interval $t_r \in [0, W_c] \cap \mathbb{Z}^+$, where W_c is the maximum backoff time. If received, MSG 2 includes uplink grant information, that indicates the RB in which the connection request (MSG 3) should be sent. The connection request specifies the requested

---

[1] When the cell size is more than twice the distance corresponding to the maximum delay spread, the eNodeB may be able to differentiate the preamble has been activated by two or more users, but only if the users are separable in terms of the Power Delay Profile [13], [14].

service type, e.g., voice call, data transmission, measurement report, etc. In case of collision the devices receive the same MSG 2, resulting in their MSG 3s colliding in the RB.

In contrast to the collisions of MSG 1, the eNodeB is able to detect collisions of MSG 3. The eNodeB only replies to the MSG 3s that did not experience collision, by sending message MSG 4, with which the required RBs are allocated or the request is denied in case of insufficient resources. The latter is however unlikely in the case of MTC, due to the small payloads. If the MSG 4 is not received within t_{CRT} since MSG 1 was sent, the random access procedure is restarted. Finally, if a device does not successfully finish all the steps of the random access procedure within $m+1$ MSG 1 transmissions, an outage is declared.

III. MODELING THE ACCESS RESERVATION PROTOCOL

We now go to the analysis of the access reservation procedure. First, we model the *One-shot transmission* and then extend it to the *m-Retransmissions* model. The numerical results cover the complete model, as depicted in Fig. 2.

A. One-Shot Transmission Model

We are interested in characterizing how often a transmission from a UE fails. This happens when the transmission is not successful in both the preamble contention and AG phases, i.e., a request from the UE must not experience a preamble collision and the uplink grant must not become stale and dropped. We model this as two independent events:

$$p_{\text{f}}(\lambda_{\text{T}}) = 1 - \left(1 - p_{\text{c}}(\lambda_{\text{T}})\right)\left(1 - p_{\text{e}}(\lambda_{\text{A}})\right), \quad (1)$$

where $p_{\text{c}}(\lambda_{\text{T}})$ is the collision probability in the preamble contention phase given UE request rate λ_{T}, while $p_{\text{e}}(\lambda_{\text{A}})$ is the probability of the uplink grant being dropped from the AG queue given preamble activation rate λ_{A}.

1) Preamble Contention Phase: We start by computing $p_{\text{c}}(\lambda_{\text{T}})$. Let d denote the number of available preambles ($d = 54$). Let the probability of not selecting the same preamble as one other UE be $1 - \frac{1}{d}$. Then the probability of a UE selecting a preamble that has been selected by at least one other UE, given N_{T} contending UEs, is:

$$\mathbb{P}\left(\text{Collision}|N_{\text{T}}\right) = 1 - \left(1 - \frac{1}{d}\right)^{N_{\text{T}}-1}. \quad (2)$$

Assuming Poisson arrivals with rate λ_T, then:

$$p_{\text{c}}(\lambda_{\text{T}}) = \sum_{i=1}^{+\infty}\left[1 - \left(1 - \frac{1}{d}\right)^{i-1} \cdot \mathbb{P}(N_{\text{T}}=i, \lambda_{\text{T}} \cdot \delta_{\text{RAO}})\right] \quad (3)$$

$$\leq 1 - \left(1 - \frac{1}{d}\right)^{\lambda_{\text{T}} \cdot \delta_{\text{RAO}}-1},$$

where $\mathbb{P}(N_{\text{T}} = i, \lambda_{\text{T}} \cdot \delta_{\text{RAO}})$ is the probability mass function of the Poisson distribution with arrival rate $\lambda_{\text{T}} \cdot \delta_{\text{RAO}}$. The inequality comes from applying Jensen's inequality to the concave function $1 - (1 - 1/d)^x$, where λ_{T} is the total arrival rate (including retransmissions), and δ_{RAO} is the average number of subframes between RAOs.[2] The computed $p_{\text{c}}(\lambda_{\text{T}})$ is thus

[2]E.g., $\delta_{\text{RAO}} = 1$ if 10 RAOs per frame and $\delta_{\text{RAO}} = 5$ if 2 RAOs per frame.

an upper bound on the collision probability.

2) Access Granting Phase: The mean number of activated preambles in the contention phase per RAO, is given by λ_{A}. As discussed in Section II, we assume that the eNodeB is unable to discern between preambles that have been activated by a single user and multiple users, respectively. This will lead to a higher λ_{A}, than in the case where the eNodeB is able to detect the preamble collisions. The main impact of this assumption is that there will be an increased rate of AG requests, even though part of these correspond to collided preambles, which even if accepted will lead to retransmissions.

The λ_{A} can be well approximated, while assuming that the selection of each preamble by the contending users is independent, by,

$$\lambda_{\text{A}} = [1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1)] \cdot d, \quad (4)$$

where $\mathbb{P}(X = k)$ is the probability of k successes, which can be well approximated with a Poisson distribution with arrival rate λ_{T}/d, i.e.:

$$\mathbb{P}(X = k) \approx \frac{(\lambda_{\text{T}}/d)^k e^{-\lambda_{\text{T}}/d}}{k!}. \quad (5)$$

To compute the outage probability due to the limitation in the AG phase, i.e., due to the maximum number of uplink grants per subframe and a maximum waiting time of t_{RAR} subframes, we consider that this subsystem can be modeled as a queuing system. We assume that the loss probability $p_{\text{e}}(\lambda_{\text{A}})$ can be seen as the *long-run fraction of costumers that are lost in a queuing system with impatient costumers* [15].

In LTE, pending uplink grants are served with a deterministic time interval (1 subframe) between each serving slot [8]. A straightforward approach would be to use an M/D/1 model structure, as presented in [15], in order to compute the drop probability. Unfortunately, the expression to compute $p_{\text{e}}(\lambda_{\text{A}})$ for the M/D/1 queue does not have a closed-form solution. However the equivalent expression for the M/M/1 queue in [15] has a closed-form solution, see eq. (6). We have compared the results of the two model types and found no noticeable difference in the computed outage numbers in practice. Thus, in the following we use the M/M/1 model to compute $p_{\text{e}}(\lambda_{\text{A}})$:

$$p_{\text{e}}(\lambda_{\text{A}}) = \frac{(1 - \rho) \cdot \rho \cdot \Omega}{1 - \rho^2 \cdot \Omega}, \text{ with } \Omega = e^{-\mu \cdot (1-\rho) \cdot \tau_{\text{q}}}. \quad (6)$$

where $\rho = \frac{\lambda_{\text{A}}}{\mu}$ is the queue load, μ is the number of uplink grants per RAR ($\mu = 3$), with $\tau_{\text{q}} = T_{\text{d}} - \frac{1}{\mu}$ and T_{d} is the max waiting time (in terms of requests) in the uplink grant queue, i.e., $T_{\text{d}} = \mu \cdot t_{\text{RAR}}$.

The fact that we are using an M/M/1 model instead of an M/D/1 model, may cause a discrepancy between the simulation and model results when the queue becomes congested ($\rho > 1$). However, we are interested in the switching point ($\rho = 1$) from which we then estimate accurately the outage breaking point, as shown in the results in section IV.
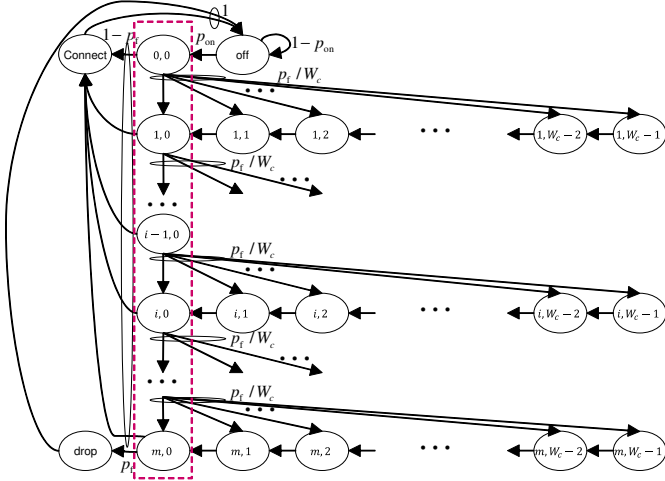
Fig. 3. Markov Chain backoff model to estimate number of required transmissions. The states in the red dashed box are used to calculate N_{TX}.

B. m-Retransmissions Model

When UEs are allowed to make retransmissions the probability of an UE becoming in outage is the probability that none of the allowed $m+1$ transmissions attempts are successful.

When retransmissions are allowed ($m > 0$), the total arrival rate λ_{T} must include the extra arrivals caused by the UEs' retransmissions. The number of retransmissions λ_{R} is however a result of the limit m and transmission error probability p_{f}, which in turn depends on the number of retransmissions λ_{R}. This chicken and egg problem can be solved iteratively using a derivative of the Bianchi model [16] applied to our system model. Specifically, we are using a model adapted to LTE, with a structure similar to the one presented in [17]. The following derivations of the number of transmissions and outage probabilities have, to the best of our knowledge, not been presented previously.

The mean number of required transmissions N_{TX} and outage probability P_{outage}, are computed with help of the Markov chain model depicted in Fig. 3. In the Markov chain model, the state index $\{i, k\}$ denotes the ith transmission attempt stage and kth backoff counter. The number of allowed retransmissions is then given by m.

Whenever the one-shot transmission is successful, depicted in Fig. 2(a), the UE enters the *connect* state:

$$\mathbb{P}\left(\text{connect}|\, i, 0\right) = 1 - p_{\text{f}}, \quad 0 \leq i \leq m.$$

Where, p_{f} is short for $p_{\text{f}}(\lambda_{\text{T}})$. Whenever the one-shot access fails, the UE increases the backoff counter:

$$\mathbb{P}\left(i, k|\, i - 1, 0\right) = \frac{p_{\text{f}}}{W_c},$$

where $0 \leq k \leq W_c - 1$ and $1 \leq i \leq m$.

At the last stage of the Markov chain, the UE enters the *drop* state if the transmission fails:

$$\mathbb{P}\left(\text{drop}|\, m, 0\right) = p_{\text{f}}(\lambda_{\text{T}}).$$

The UE enters the *off* state after the *connect* or the *drop* states, with probability:

$$\mathbb{P}\left(\text{off}|\, \text{drop}\right) = \mathbb{P}\left(\text{off}|\, \text{connect}\right) = 1.$$

From the *off* state, the node enters the first transmission state $\{0, 0\}$ with probability p_{on}:

$$\mathbb{P}\left(0, 0|\, \text{off}\right) = p_{\text{on}}.$$

where the probability p_{on} is defined as $p_{\text{on}} = 1 - e^{-\lambda_{\text{I}}}$.

Let $b_{i,k}$ be the steady state probability that a UE is at state $\{i, k\}$. Then $b_{i,k}$ can be derived as:

$$b_{i,k} = \frac{W_{\text{c}} - k}{W_{\text{c}}} p_{\text{f}} b_{i-1,0} = \frac{W_{\text{c}} - k}{W_{\text{c}}} p_{\text{f}}^i b_{0,0} = \frac{W_{\text{c}} - k}{W_{\text{c}}} p_{\text{f}}^i p_{\text{on}} b_{\text{off}}, \quad (7)$$

for $1 \leq i \leq m$ and $0 \leq k \leq W_{\text{c}} - 1$.

Let b_{connect} be the steady state probability that a node is at *connect* state:

$$
\begin{aligned}
b_{\text{connect}} &= \sum_{i=0}^{m} \left(1 - p_{\text{f}}\right) b_{i,0} = \sum_{i=0}^{m} \left(1 - p_{\text{f}}\right) p_{\text{f}}^i p_{\text{on}} b_{\text{off}} \\
&= \left(1 - p_{\text{f}}^{m+1}\right) p_{\text{on}} b_{\text{off}}.
\end{aligned}
$$

By imposing the probability normalization condition, as detailed in Appendix A, we find b_{off} as:

$$b_{\text{off}} = \frac{2\left(1 - p_{\text{f}}\right)}{2\left(1 - p_{\text{f}}\right)\left(1 + 2 p_{\text{on}}\right) + p_{\text{on}}\left(W_{\text{c}} + 1\right) p_{\text{f}}\left(1 - p_{\text{f}}^m\right)}.$$

Since a transmission will eventually either finish successfully in the *connect* state or unsuccessfully in the *drop* state, the outage probability can be computed as:

$$P_{\text{outage}} = \frac{b_{\text{drop}}}{b_{\text{drop}} + b_{\text{connect}}} = p_{\text{f}}^{m+1}, \quad (8)$$

where b_{drop} and b_{connect}, whose derivations are shown in the Appendix A, can be computed as:

$$b_{\text{connect}} = \frac{2\left(1 - p_{\text{f}}\right)\left(1 - p_{\text{f}}^{m+1}\right) p_{\text{on}}}{2\left(1 - p_{\text{f}}\right)\left(1 + 2 p_{\text{on}}\right) + p_{\text{on}}\left(W_{\text{c}} + 1\right) p_{\text{f}}\left(1 - p_{\text{f}}^m\right)}, \quad (9)$$

$$b_{\text{drop}} = \frac{2\left(1 - p_{\text{f}}\right) p_{\text{f}}^{m+1} p_{\text{on}}}{2\left(1 - p_{\text{f}}\right)\left(1 + 2 p_{\text{on}}\right) + p_{\text{on}}\left(W_{\text{c}} + 1\right) p_{\text{f}}\left(1 - p_{\text{f}}^m\right)}. \quad (10)$$

The number of required transmissions can be estimated from the steady state probabilities, keeping in mind that $b_{i,0}/b_{0,0}$ represents the probability of using $i + 1$ or more transmission attempts to deliver a packet, and $b_{m,0}/b_{0,0}$ is the probability of using exactly $m+1$ transmission attempts:

$$
\begin{aligned}
N_{\text{TX}}(\lambda_{\text{T}}) &= \frac{\left(\sum_{i=0}^{m-1} (i+1) \cdot (b_{i,0} - b_{i+1,0})\right) + (m+1) \cdot b_{m,0}}{b_{0,0}} \\
&= (1 - p_{\text{f}}) \sum_{i=0}^{m-1} (i+1) p_{\text{f}}^i + (m+1) p_{\text{f}}^m = \frac{1 - p_{\text{f}}^{m+1}}{1 - p_{\text{f}}}.
\end{aligned}
\tag{11}
$$

From the number of transmissions, the value of λ_{T} can be solved iteratively using the fixed point equation:

$$\lambda_{\text{T}} = N_{\text{TX}}(\lambda_{\text{T}}) \cdot \lambda_{\text{I}} = \lambda_{\text{I}} \frac{1 - p_{\text{f}}(\lambda_{\text{T}})^{m+1}}{1 - p_{\text{f}}(\lambda_{\text{T}})}. \quad (12)$$

(a) Outage probability, $m = 0$ (b) Expected number of transmissions, $m = 9$ (c) Outage probability, $m = 9$

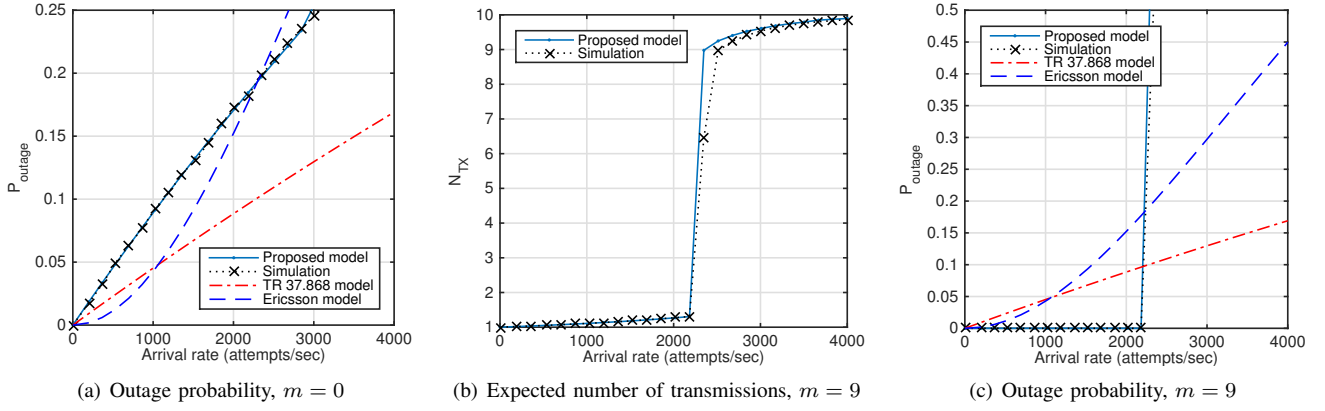Fig. 4. Plots for RACH configuration with 2 RAOs per frame ($\delta_{\mathrm{RAO}} = 5$). Ericsson model refers to [6].

(a) Outage probability, $m = 0$ (b) Expected number of transmissions, $m = 9$ (c) Outage probability, $m = 9$
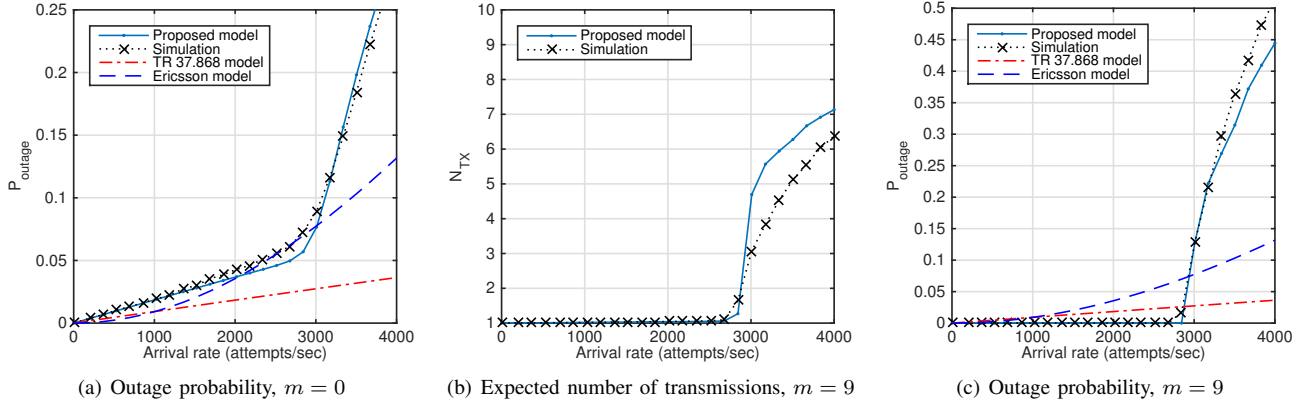
Fig. 5. Plots for RACH configuration with 10 RAOs per frame ($\delta_{\mathrm{RAO}} = 1$). Ericsson model refers to [6].

For the results presented in Sec. IV we found that less than 20 iterations were needed to reach convergence (less than 1% change between consecutive iterations).

IV. NUMERICAL RESULTS

In our evaluation, we consider two PRACH configurations, namely the typical configuration with 5 subframes between every RAO [18] and the configuration with one RAO every subframe. Further, we consider first the case where only a single transmission is allowed (one-shot, $m=0$) and then the more realistic configuration of $m=9$ allowed retransmissions. The model results are compared with a simulator that implements the full LTE access reservation protocol as defined in [11] and [12] given parameters in Table I.

Parameter	Value
Preambles per RAO (d)	54
Subframes between RAOs (δ_{RAO})	1 or 5
Max number of retransmissions (m)	0 or 9
Uplink grants per RAR (μ)	3
System bandwidth	5 MHz
eNodeB processing time	3 ms
MSG 2 window (t_{RAR})	5 ms or 10 ms
Contention time-out (t_{CRT})	48 ms
Backoff limit (W_{c})	20 ms
UE processing time	3 ms

TABLE I
LTE SIMULATION AND MODEL PARAMETERS

A. One-shot Transmission ($m=0$)

In Fig. 4(a) and 5(a) the outage probabilities are depicted for $m = 0$. There, the proposed model has a much better fit to the simulation results than the 3GPP TR 37.868 model [4, Sec. B.1] and the Ericsson model in [6, eq. (6)]. Specifically, in Fig. 4(a) where the preamble collisions are the main error cause, the TR 37.868 and Ericsson models are much worse than the proposed model. From Fig. 5(a) it is clear that those models are not accounting for the limited number of uplink grants per RAR that starts to have an impact around $\lambda_{\mathrm{I}} = 2700$ attempts/sec, causing an upward bend in the outage curve.

B. $m=9$ Retransmissions

In the typical configuration where retransmissions are allowed, a necessary feature of our model is that it is able to account for the feedback impact of retransmissions on the arrival rate λ_{T}. An intermediate metric that allows to study this is the number of transmissions per new data packet N_{TX}. This is shown in Figs. 4(b) and 5(b). In Fig. 4(b) the number of transmissions is estimated accurately leading to a well-fitting estimation of the outage in 4(c). For the case of 10 RAOs per frame, the Markov chain model slightly overestimates the number of transmissions. However, the breaking points in the curves are the same, meaning that the supported arrival rate

in the simulation on Fig. 5(c) is closely matched by the one in the model.

Finally, the results show that the proposed model is superior to the existing models from the literature, as they do not capture the feedback impact of the retransmissions and are therefore not able to estimate the system outage capacity.

The presented results also reveal an interesting insight in dimensioning the LTE access reservation parameters. Given that there is a 5 times difference in resource usage for RAOs (2 vs 10 RAOs per frame), the gain in supported arrival rate λ_I is quite modest, increasing from around $\lambda_I = 2250$ to around $\lambda_I = 2800$, i.e., a 25% increase. In order to further increase the capacity of the system, it is necessary to simultaneously increase the number of RARs per subframe.

V. Conclusions and Outlook

In this paper we have presented a low-complexity, yet accurate model to estimate the outage capacity of the LTE access reservation protocol for machine-type communications, where the small payload sizes mean that system resources are typically not the limiting factor. The model accounts for both contention preamble collisions and the limited number of uplink grants in the random access response message, as well as the feedback impact that the resulting retransmissions has on the random access load. For the considered typical LTE configurations, the model is able to very accurately estimate the system outage capacity. This puts it forward as a useful tool in system dimensioning, as it allows to replace time-consuming simulations with click-speed calculations.

Future work should look into how diverse channel conditions and diverse traffic patterns of users can be efficiently included in the model. While the outage metric is very important from a planning perspective, other metrics such as access delay or transmission time would be very relevant to be able to estimate accurately when considering real-time machine-type communications.

Acknowledgment

Appendix

The following steps are taken to derive b_{off} by imposing the probability normalization condition:

$$1 = b_{\text{off}} + b_{\text{conn}} + b_{\text{drop}} + b_{0,0} + \sum_{i=1}^{m}\sum_{k=0}^{W_c-1} b_{i,k}$$

$$= b_{\text{off}} + 2p_{\text{on}}b_{\text{off}} + \sum_{i=1}^{m}\sum_{k=0}^{W_c-1} \frac{W_c - k}{W_c} p_f^i p_{\text{on}} b_{\text{off}}$$

$$= b_{\text{off}} + 2p_{\text{on}}b_{\text{off}} + p_{\text{on}}b_{\text{off}}\left(\frac{W_c+1}{2}\right) p_f \frac{1-p_f^m}{1-p_f}.$$

The derivation of b_{connect} is as follows:

$$b_{\text{connect}} = 1 - b_{\text{off}} - p_{\text{on}}b_{\text{off}}(p_f^{m+1} - 1 - \left(\frac{W_c+1}{2}\right) p_f \frac{1-p_f^m}{1-p_f})$$

$$= \frac{2(1-p_f)\left(1-p_f^{m+1}\right)p_{\text{on}}}{2(1-p_f)(1+2p_{\text{on}})+p_{\text{on}}(W_c+1)\,p_f\,(1-p_f^m)}. \quad (13)$$

The derivation of b_{drop} is as follows:

$$b_{\text{off}} = (1 - p_{\text{on}})\,b_{\text{off}} + b_{\text{drop}} + b_{\text{connect}}$$

$$b_{\text{drop}} = p_{\text{on}}b_{\text{off}} - b_{\text{connect}}$$

$$= \frac{2\left(1-p_f\right)p_f^{m+1}p_{\text{on}}}{2(1-p_f)(1+2p_{\text{on}})+p_{\text{on}}(W_c+1)\,p_f\left(1-p_f^m\right)}. \quad (14)$$

References

[1] G. Corrales Madueño, C. Stefanovic, and P. Popovski, "Reengineering GSM/GPRS Towards a Dedicated Network for Massive Smart Metering," in *Proc. of the IEEE Internation Conference on Smart Grid Communications (SmartGridComm 2014)*, Nov. 2014.

[2] 3GPP, "R1-061369: LTE random-access capacity and collision probability," Tech. Rep. RAN1#45, May 2006.

[3] ——, "R2-112198: Clarification on the discussion of RACH Collision Probability," Tech. Rep. RAN2#73bis, April 2011.

[4] ——, "TR 37.868: Study on RAN improvements for machine-type communications, rel. 11," Tech. Rep., September 2011.

[5] R.-G. Cheng, C.-H. Wei, S.-L. Tsao, and F.-C. Ren, "RACH Collision Probability for Machine-Type Communications," in *Proc. of the IEEE Vehicular Technology Conference (VTC Spring 2012)*, May 2012.

[6] C. Ubeda, S. Pedraza, M. Regueira, and J. Romero, "LTE FDD physical random access channel dimensioning and planning," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall 2012)*, Sept. 2012.

[7] C. Karupongsiri, K. S. Munasinghe, and A. Jamalipour, "Random access issues for smart grid communication in LTE networks," in *Proc. of the International Conference on Signal Processing and Communication Systems (ICSPCS 2014)*, Dec. 2014.

[8] B. Yang, G. Zhu, W. Wu, and Y. Gao, "M2M Access Performance in LTE-A System," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 3–10, Jan. 2014.

[9] P. Osti, P. Lassila, S. Aalto, A. Larmo, and T. Tirronen, "Analysis of PDCCH performance for M2M traffic in LTE," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4357–4371, Nov. 2014.

[10] 3GPP, "TS 36.201 E-UTRA LTE physical layer; general description," Tech. Rep., 2015.

[11] ——, "TS 36.321 E-UTRA medium access control (MAC) protocol specification," Tech. Rep., 2015.

[12] ——, "TS 36.213 E-UTRA physical layer procedures," Tech. Rep., 2015.

[13] S. Sesia, I. Toufik, and M. Baker, *LTE-The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 2011.

[14] H. Thomsen, N. Pratas, C. Stefanovic, and P. Popovski, "Analysis of the LTE Access Reservation Protocol for Real-Time Traffic," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1616–1619, Aug. 2013.

[15] A. G. De Kok and H. Tijms, "A queueing system with impatient customers," *Journal of Applied Probability*, vol. 22, no. 3, pp. 688–696, Sept. 1985.

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

[17] X. Yang, A. Fapojuwo, and E. Egbogah, "Performance Analysis and Parameter Optimization of Random Access Backoff Algorithm in LTE," in *Proc. of the IEEE Vehicular Technology Conference (VTC Fall 2012)*, Sept. 2012.

[18] 3GPP, "MTC simulation assumptions for RACH performance evaluation," 3rd Generation Partnership Project (3GPP), TR R2-105212, Aug. 2010.