Aalborg Universitet

AALBORG UNIVERSITY

Analysis of the LTE Access Reservation Protocol for Real-Time Traffic

Thomsen, Henning; Kiilerich Pratas, Nuno; Stefanovic, Cedomir; Popovski, Petar

Published in:
I E E E Communications Letters

DOI (link to publication from Publisher):
10.1109/LCOMM.2013.060513.130995

Publication date:
2013

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Thomsen, H., Kiilerich Pratas, N., Stefanovic, C., & Popovski, P. (2013). Analysis of the LTE Access Reservation Protocol for Real-Time Traffic. *I E E E Communications Letters*, *17*(8), 1616 - 1619. https://doi.org/10.1109/LCOMM.2013.060513.130995

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 05, 2025

Analysis of the LTE Access Reservation Protocol for Real-Time Traffic

Henning Thomsen, Nuno K. Pratas, Čedomir Stefanović, Petar Popovski

***Abstract*—LTE is increasingly seen as a system for serving real-time Machine-to-Machine (M2M) communication needs. The asynchronous M2M user access in LTE is obtained through a two-phase access reservation protocol (*contention* and *data* phase). Existing analysis related to these protocols is based on the following assumptions: (1) there are sufficient resources in the *data phase* for all detected contention tokens, and (2) the base station is able to detect collisions, i.e., tokens activated by multiple users. These assumptions are not always applicable to LTE - specifically, (1) due to the variable amount of available data resources caused by variable load, and (2) detection of collisions in contention phase may not be possible. All of this affects transmission of real-time M2M traffic, where data packets have to be sent within a deadline and may have only one contention opportunity. We analyze the features of the two-phase LTE reservation protocol and asses its performance, when assumptions (1) and (2) do not hold.**

***Index Terms*—Access Reservation Protocols, LTE, M2M communications**

I. Introduction

An access reservation protocol is instrumental in any multi-user communication system in order to enable users to connect asynchronously or transmit intermittently [1]. The Long Term Evolution (LTE) system [2] uses an access protocol consisting of two phases: a *contention phase*, where each user contends by activating a particular reservation token chosen from the set of available tokens; and a *data phase*, where the reservation tokens (i.e., token holders) detected by the base station (BS) get assigned resources for the data transfer. The asynchronous access in LTE gains importance as the needs to support traffic related to Machine-to-Machine (M2M) communications gets increasingly important. In many cases, M2M traffic is a real-time traffic, where data packets become obsolete after a deadline and thus may undergo only a single contention and data phases, i.e., unsuccessful transmissions cannot be postponed for later contention or scheduled to a later data phase.

The available analysis of the two-phase access reservation protocols typically assumes that: (1) there are sufficient resources in the data phase to serve all detected reservation tokens; (2) the BS is able to discern between reservation tokens activated by one or more than one users, i.e., the contention phase has a ternary output (idle, single or collision). However, assumption (1) does not hold in cellular networks such as LTE, where the data phase has limited number of resources, while the network load is variable; this implies that there is a possibility that the users with real-time traffic that contended successfully may not be assigned a data transmission slot at all [2]. Assumption (2), by default does not hold in LTE, as the BS may not able to discern if a token was activated by one or multiple users [3, Sec. 17.5.2.3]. In other words, there are practical setups in which the BS can "see" that a preamble has been activated, but it does not know how many users activated it. This implies that in the contention phase, collisions "over" tokens are treated as singles, i.e., the output of the contention phase is binary (idle or active) instead of the commonly assumed ternary output (idle, single or collision).

In LTE, a reservation token is activated by transmitting a specific preamble in the random access sub-frame [4]. The preambles are chosen from the orthogonal set of preambles obtained from Zadoff-Chu sequences [5]. Due to orthogonality of the preambles, the LTE contention phase can be modeled as a framed slotted ALOHA scheme, where "slots" represent preambles over which the users contend.

Asynchronous M2M communications based on LTE are considered in several works found in the literature. In [6], the authors propose a preamble retransmission method, subject to optimization of the transmission rate. In [7], a packet aggregation method is proposed. Here, M2M devices do not necessarily transmit their packets immediately, but buffer them until a certain threshold. A closed form expression of the collision probability of M2M traffic at the LTE contention phase is provided in [8]. A recent work using a combinatorial model to study the random access in LTE is given in [9]. Different from [9], we also consider the data phase in our analysis.

An early study of the ALOHA protocol in a reservation framework was done in [10], and [11] consideres two reservation methods based on framed ALOHA. Access reservation protocols with several parallel data channels have been studied in [12]. Herein, the authors find the optimal ratio of control to data channels, and the optimal number of data channels, in terms of throughput. Note that in contrast to [12], in this letter, we consider collisions in the contention phase to be non-destructive. Finally, we point out a recent analysis of random access protocols in the context of RFID systems given in [13]; the results presented therein are directly applicable to the contention phase of the access reservation protocol and

The authors are with the Department of Electronic Systems, Aalborg University, Aalborg, Denmark (e-mail: {ht, nup, cs, petarp}@es.aau.dk).

The research presented in this paper was partly 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" and performed partly in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

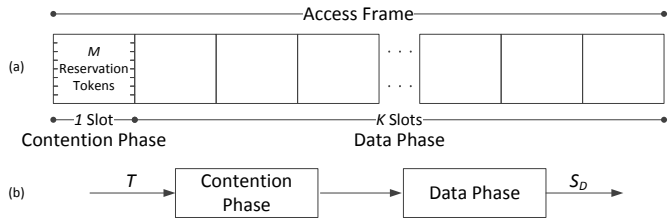

Fig. 1. (a) Access Frame and (b) System Model.

Variable	Description
M	No. of reservation tokens
K	No. of TDM slots
T	No. of accessing users
S_D	No. of users succeeding in the data phase
s	No. of reservation tokens selected by one user
c	No. of reservation tokens selected by more than one user
k	$\min\{s+c, K\}$
m	$\min\{M, T\}$
σ	Success Rate
ρ	Efficiency

TABLE I
DEFINITION OF USED VARIABLES.

are used as a starting point for the analysis presented in the paper.

In this letter, we derive the exact probability mass function (pmf) that a number of reservation tokens activated by a single user are assigned resources in the data phase, when assumptions (1) and (2) do not hold. Based on the obtained results, we calculate the corresponding one-shot success rate and efficiency of the LTE access reservation scheme. The presented analysis and derived results are directly applicable to the LTE access reservation scheme for the increasingly important case of asynchronously served users with real-time constraints.

The remainder of this paper is organized as follows. In Section II, we present the system model. Section III elaborates the method to obtain the exact pmf of the number of reservation tokens that are activated by a single user and that are assigned resources in the data phase, following by the derivation of the success rate and efficiency. Examples demonstrating derived results are given in Section IV, while the letter is concluded in Section V.

II. System Model

Fig. 1(a) shows a simplified version of the LTE access reservation protocol that captures the details essential for the presented analysis [4]. The access reservation is composed of a contention phase and data phase. The contention phase lasts a single slot and is modeled as a variant of a framed slotted ALOHA, where users contend over a set of of available tokens[1]; we assume that in this slot, there are available M reservation tokens for contention. The data phase is a actually a Time Division Multiplexing (TDM) scheme and we assume that there are available K resource slots (i.e., K TDM slots). The combination of the two phases is denoted as an access frame, consisting of $K+1$ slots in total. Finally, we assume that there are T users contending for the available resource slots.

The access reservation protocol operates as follows:

1) Each of T users activates randomly and independently one of the M available tokens. A token can be activated by more than one user.
2) Base Station (BS) detects all activated tokens, irrespective whether they have been activated by one or several users [3, Sec. 17.5.2.3]. The BS chooses uniformly randomly K tokens from the set of detected tokens.
3) The selected tokens are assigned a resource slot each and the corresponding users, i.e., token holders, are informed about the respectively assigned slots through the feedback channel.
4) The selected token holders transmit their data packets in the assigned resource slots. If two or more holders activated the same token and thus were assigned the same resource slot, their transmissions collide and are considered as lost.

The assumption that the BS is unable to detect collision in contention phase holds in small cells [3, Sec. 17.5.2.3], and refers to the worst case scenario where the detected preamble does not reveal anything about the number of users that transmitted it[2]. Obviously, if the BS knows that there are two or more users using a certain preamble, then a straightforward way to operate is not to assign any resource slot to the preamble, thus preventing collisions and the respective resource waste in the data phase.

III. Analysis

A. Derivation of the pmf

The conditional probability of having assigned $S_D = d$ contention tokens used exactly by one user each from the contention phase of size M to the data phase of size K when $T = t$, denoted by $P(S_D = d \mid T = t, K, M)$, is derived in this section. The parameters used in the derivation are listed in Table I.

Suppose that out of the M available tokens, s are used by exactly one user each (singles), and c are used by two or more users (collisions), where $0 \leq s+c \leq M$. From the M available tokens, s single tokens can be selected in $\binom{M}{s}$ ways.[3] Further, as the tokens are distinguishable, the number of ways in which s of the T users can be selected is $T \cdot (T-1) \cdots (T-s+1)$ ways.

From the remaining $M-s$ tokens, we choose c for the colliding users, which can be done in $\binom{M-s}{c}$ ways. As the tokens are distinguishable, we need to count all permutations of them, which equals $c!$. Further, the number of ways in which

[1] In contrast to standard ALOHA, in the considered model there are no collisions, as elaborated in the letter.

[2] Note that, in practical LTE systems, if the cell size is more than twice the distance corresponding to the maximum delay spread, the BS may, in some circumstances, be able to differentiate the transmission of the same preamble by two or more users, provided that the users are separable in terms of the Power Delay Profile [3]. However, the analysis of such operation is straightforward and therefore not of interest in this letter.

[3] We define $\binom{n}{k} = 0$ when $n < k$.

$$P(S_D = d \mid T = t, K, M) = \sum_{0 \le s \le m} \sum_{0 \le c \le m-s} \frac{\binom{M}{s} T \cdot (T-1) \cdots (T-s+1) \binom{M-s}{c} S_2(T-s, c) c! \; \binom{s}{d}\binom{c}{k-d}}{M^T \qquad \binom{s+c}{k}} \tag{1}$$

$T - s$ users can be distributed among c tokens such that there are at least two users selecting each token is given by the 2-associated Stirling number of the second kind $S_2(T-s, c)$ [14, pp.221-222], which can be computed using the recurrence relation (7) exposed in the Appendix.

The total number of ways in which T users can select among M tokens without restriction is M^T. Therefore, the probability of the T users selecting among M tokens such that there are s tokens used by exactly one user each, and c tokens used by two or more users each, equals

$$\frac{\binom{M}{s} T \cdot (T-1) \cdots (T-s+1) \binom{M-s}{c} S_2(T-s, c) c!}{M^T}, \tag{2}$$

as similarly derived in [13], although in a different context.

In the data phase, the $s+c$ used tokens are mapped randomly to the K resource slots. Here we distinguish between two cases, $s + c \le K$ and $s + c > K$. In the first case, all used tokens are assigned to data slots. In the second case, K out of the $s + c$ tokens are randomly selected and assigned to the data slots. We then have that the probability of selecting d slots out of the K, such that the selected slots are used by one user each, is given by the hypergeometric distribution

$$\frac{\binom{s}{d}\binom{c}{k-d}}{\binom{s+c}{k}}. \tag{3}$$

where $k = \min\{s + c, K\}$. Note that when $k = s + c$, then Eq. (3) equals 1 when $s = d$, and 0 otherwise.

The probability of selecting d slots containing one user each, given that $T = t$ users transmit, is obtained by summing over all cases of s and c such that $0 \le s+c \le M$, which yields the complete expression of $P(S_D = d \mid T = t, K, M)$ presented in Eq. (1), where $m = \min\{M, T\}$ and $k = \min\{s + c, K\}$.

We conclude by noting that (1) holds when $M > K$ and also when $M \le K$.

B. Success Rate and Efficiency

We use the pmf derived in the previous subsection in obtaining an expression for the success rate, defined as the expected value of S_D given a number of transmitting users T

$$\sigma = \frac{E\left[S_D | T = t, K, M\right]}{t} = \frac{1}{t} \sum_{s=0}^{\min\{K,M\}} s \cdot P(S_D = s | T = t). \tag{4}$$

As defined, the success rate takes into account both phases of the LTE access reservation scheme and measures the expected fraction of successfully accessing users.

In order to assess how well the slots of the access frame (see Fig. 1(a)) are utilized, we define the efficiency, given $T = t$, K and M, as

$$\rho = \frac{E\left[S_D | T = t, K, M\right]}{K + 1} \tag{5}$$

where the denominator corresponds to the length of the access frame in slots.

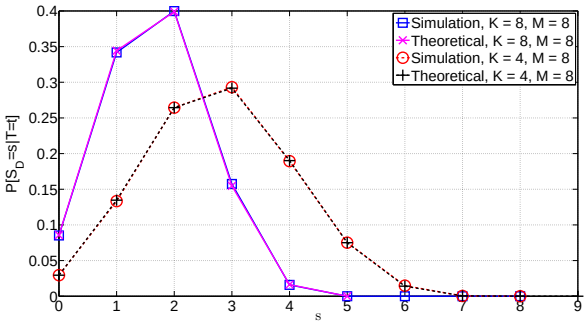

Fig. 2. Comparison of theoretical and simulated pmfs for $M > K$ and $M \le K$.

IV. Results

In this section, we give examples of the pmf in Eq.(1), the success rate in Eq.(4) and efficiency in Eq.(5), both for the cases when $M > K$ and $M \le K$.

A. pmf Evaluation

We first describe the method to obtain the pmf from the simulation. The method can be summarized in the following three steps.

1) Let the T users select their tokens randomly and independently. For every token, count the number of times this token is selected by the users.
2) To simulate a limited data phase, select a random subset of size K from the set of tokens. Count the number of successful tokens in this subset, i.e. tokens used by one user. Call this number S.
3) Let $S^{(n)}$ be the value of S obtained in iteration $1 \le n \le N$, where N is the total number of iterations. The simulated pmf is then

$$\widehat{P}(S_D = s \mid T = t) = \frac{\sum_{n=1}^{N} \mathbb{I}(S^{(n)})}{N}, \tag{6}$$

where $\mathbb{I}(\cdot)$ is an indicator function, giving $\mathbb{I}(S^{(n)}) = s$ if $S^{(n)} = s$, and 0 otherwise.

Fig. 2 compares the analytical and simulated pmfs for $T = 12$ accessing users, when number of tokens is set to $M = 8$ and the number of resource slots is $K = 4$ and $K = 8$. The analytical pmf is obtained from Eq. (1), while the simulated one is obtained from running $N = 100000$ simulation iterations. We observe a correlation between the analytical and simulated pmfs, validating the presented analysis.

B. Success Rate and Efficiency Evaluation

Fig. 3 and Fig. 4 show respectively the success rate σ, Eq. (4), and the efficiency ρ, Eq. (5) as functions of number of users T, for $M = 8$ and varying K.

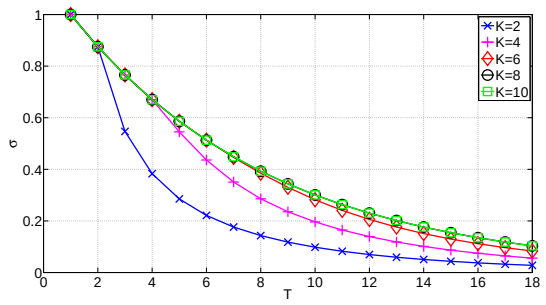

Fig. 3. Comparison of success rate σ, for $M = 8$, varying user load T, and different sizes of the data phase K.

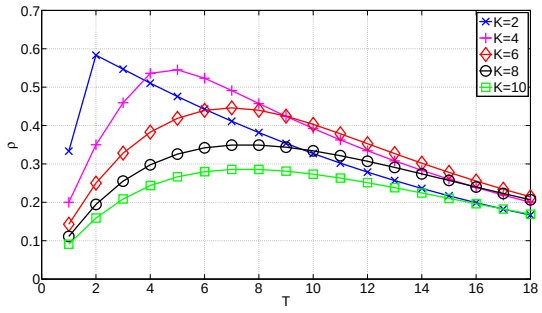

Fig. 4. Comparison of efficiency ρ, for $M = 8$, varying user load T, and different sizes of the data phase K.

From Fig. 3 it can be observed that success rate decreases as the number of users T increase; this is due to increasing probability of multiple users selecting the same tokens and, consequently, increasing number of collisions happening in the data phase. On the other hand, increasing the number of available resource slots K increases the success rate at first, until K reaches the number of available tokens M. Afterwards, there is no benefit in increasing K, as no more than M users can be successfully detected and admitted in the system and $K - M$ resource slots will always be left unassigned.

From Fig. 4 it is clear that the higher efficiency is achieved when is K considerably lower than M, which is not a straightforward conclusion. Again, this is because of multiple users selecting the same preambles, which results both in *collision* and *idle* slots in the data phase; the chances for the latter increase with the number of available resource slots K. We end this section by noting that, by using the framework presented in paper, the optimal number of resource slots K that maximizes the efficiency can be calculated, for fixed number of tokens M and number of users T.

V. Conclusion

In this letter we studied a LTE based access reservation protocol and provided a method to obtain the exact pmf describing the number of reservation tokens activated by a single user that gets assigned resources in the data phase. The obtained results are applicable to the case where there are not enough resources in the data phase to serve all detected reservation tokens, and when the base station is not able to discern between reservation tokens selected by one or more than one user; both assumptions may occur in practice, affecting the operation of the access reservation protocol.

Further, based on the presented method we derived the one-shot success rate and efficiency of the scheme, which can be used as performance measures of the constrained access reservation systems, such as LTE, in the emerging scenarios with real-time M2M communication, when there is a limited time to carry out the contention and the data transmissions. Finally, it was observed that although the success rate is maximized when there are the same number of contention and data resources, there is a non-negligible efficiency tradeoff for doing so.

Appendix

Here, a method to compute the 2-associated Stirling numbers of the second kind is presented, based on [14]. By definition, $S_2(n, k)$ is the number of ways in which k objects can be put into n boxes, such that each box contains at least 2 objects. The values of $S_2(n, k)$ can be computed using the recurrence relation

$$S_2(n+1, k) = kS_2(n, k) + nS_2(n-1, k-1), \qquad (7)$$

with initial condition $S_2(2, 1) = 1$. If $k > \lfloor \frac{n}{2} \rfloor$, $n \leq 0$ or $k \leq 0$, then $S_2(n, k) = 0$.

References

[1] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Prentice-Hall, 1987.

[2] *LTE physical layer; General description*, 3GPP TS 36.201.

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

[4] *Medium Access Control (MAC) protocol specification*, 3GPP TS 36.321.

[5] *Physical layer procedures*, 3GPP TS 36.213.

[6] Z. Jiang and X. Zhong, "Fast Retrial and Dynamic Access Control Algorithm for LTE-Advanced Based M2M Network," in *AICT 2012, The Eighth Adv. Internat. Conf. on Telecommun.*, 2012, pp. 24–28.

[7] K. Zhou and N. Nikaein, "Packet Aggregation for Machine Type Communications in LTE with Random Access Channel," in *IEEE Wirel. Commun. Netw. Conf., Apr. 7-10, 2013, Shanghai, China*, 04 2013. [Online]. Available: http://www.eurecom.fr/publication/3938

[8] R.-G. Cheng, C.-H. Wei, S.-L. Tsao, and F.-C. Ren, "Rach Collision Probability for Machine-type Communications," in *IEEE Veh. Tech. Conf., Spring 2012*. IEEE, 2012, pp. 1–5.

[9] C.-H. Wei, R.-G. Cheng, and S.-L. Tsao, "Modeling and Estimation of One-Shot Random Access for Finite-User Multichannel Slotted ALOHA Systems," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1196 –1199, Aug. 2012.

[10] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A System for Broadcast Communication: Reservation-ALOHA," *Proc. 6th Hawaii Int. Conf. Syst. Sci*, pp. 596–603, 1973.

[11] W. Szpankowski, "Analysis and Stability Considerations in a Reservation Multiaccess System," *IEEE Trans. Commun.*, vol. 31, no. 5, pp. 684–692, May 1983.

[12] Y. S. Han, J. Deng, and Z. J. Haas, "Analyzing Multi-Channel Medium Access Control Schemes with ALOHA reservation," *IEEE Trans. Wirel. Commun.*, vol. 5, no. 8, pp. 2143–2152, 2006.

[13] J. Alcaraz, J. Vales-Alonso, E. Egea-López, and J. Garcia-Haro, "A Stochastic Shortest Path Model to Minimize the Reading Time in DFSA-Based RFID Systems," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 341–344, 2013.

[14] L. Comtet, *Advanced Combinatorics. The Art of Finite and Infinite Expansions, Revised and Enlarged edition*. Riedel Publishing Co., Dordrecht, 1974.