Code-Expanded Random Access for Machine-Type Communications

Nuno K. Pratas, Henning Thomsen, Čedomir Stefanović, Petar Popovski
Department of Electronic Systems, Aalborg University, Denmark
Email: nup@es.aau.dk, ht1703@gmail.com and {cs,petarp}@es.aau.dk

Abstract—The random access methods used for support of machine-type communications (MTC) in current cellular standards are derivatives of traditional framed slotted ALOHA and therefore do not support high user loads efficiently. Motivated by the random access method employed in LTE, we propose a novel approach that is able to sustain a wide random access load range, while preserving the physical layer unchanged and incurring minor changes in the medium access control layer. The proposed scheme increases the amount of available contention resources, without resorting to the increase of system resources, such as contention sub-frames and preambles. This increase is accomplished by expanding the contention space to the code domain, through the creation of random access codewords. Specifically, in the proposed scheme, users perform random access by transmitting one or none of the available LTE orthogonal preambles in multiple random access sub-frames, thus creating access codewords that are used for contention. In this way, for the same number of random access sub-frames and orthogonal preambles, the amount of available contention resources is drastically increased, enabling the support of an increased number of MTC users. We present the framework and analysis of the proposed code-expanded random access method and show that our approach supports load regions that are beyond the reach of current systems.

I. INTRODUCTION

In the past few years there has been an increase in the number of networked machines on current networks designed for human-centric communications. This has led to a shift in the conventional perception on communications towards device-centric communications, which is independent of human interaction [1], [2]. While the requirements of traditional human-centric communications are high bit-rates and lower latency, in device-centric communications, the main requirement is the massive transmission of simultaneous low data rate messages.

The main challenge in Machine-to-Machine (M2M) communications, denoted as Machine-Type Communications (MTC) by 3GPP, is to adapt cellular networks to efficiently handle MTC traffic characteristics, specifically the load generated by massive simultaneous low data rate transmissions. Within 3GPP, there is an ongoing study on the adaptation of the 3GPP cellular networks to handle the MTC traffic [3], [4], through the inclusion of enhanced load control mechanisms in the Radio Access Network (RAN). This is paramount, because due to the expected large number of deployed MTC devices, the cellular networks are expected to withstand traffic bursts [5]. In such situations, radio and signalling network congestions may occur due to mass concurrent transmissions [6], which can lead to large delays, packet loss and, in the extreme case, service unavailability.

In 3GPP there have been proposed several solutions for managing the random access load [3]. These solutions include access class barring [7], [8], and back-off schemes, where the network controls the load by restricting and delaying the random access, thus spreading the random access in time and reducing the effect of batch arrivals. Other solutions divide and adapt the amount of random access resources for human-centric and device-centric traffic, such as orthogonal random access [9] and dynamic random access [2], where overload due device-centric traffic will not affect the human-centric traffic and vice versa.

In this paper we propose a random access method inspired by the LTE random access, which is in line with the dynamic Random Access Channel (RACH) resource allocation approach. Here the dynamic resource allocation is not accomplished through traditional means, such as the increase of the number of available preambles and/or random access sub-frames, but instead by introducing the concept of code-expanded random access. The motivation behind this proposal is to enable existing systems to sustain a large and bursty random access loads, while preserving the physical layer unchanged and incurring minimal changes in the medium access control layer.

To illustrate what is meant by code-expanded random access, consider the diagrams in Fig. 1 where the selection of preambles and sub-frames in the reference and in the proposed code-expanded random access schemes are depicted. In the reference scheme, each MTC user (M-UE) performs the random access by selecting one of the available preambles, as depicted in Fig. 1(a). The Base Station (BS) then discerns between M-UEs according to the observed preambles in each random access sub-frame, as shown in Fig. 1(c). In the proposed scheme, depicted in Fig. 1(b), the random access sub-frames are grouped in virtual frames and the random access is performed on the virtual frame level. In each sub-frame of the virtual frame, the M-UE transmits one or none of the available preambles (we assume that in the case when none of the preambles is transmitted, the M-UE actually transmits an idle preamble, henceforth denoted as I), thus creating an access codeword with length equal to the length of the virtual frame. In this case, the BS observes the distribution of preambles
over the virtual frame sub-frames, i.e., observes a set of access codewords that are then used to discern between M-UEs.

The proposed code-expanded random access scheme is a generalization of the reference scheme. In the reference scheme the codewords are composed of just one preamble sent in one of the sub-frames of the virtual frame, with the remaining sub-frames at idle. In the proposed scheme, each M-UE selects a codeword consisting of a randomly chosen preamble (including the idle preamble) per every sub-frame of the virtual frame. In this way, the number of contention resources is expanded and the amount of collisions is reduced - a collision occurs when two or more M-UEs select the same codeword, as depicted in Fig. 1(c).

On the other hand, while reducing the amount of collisions, the code-expanded generalization introduces a shortcoming not present in the reference scheme. In the reference scheme, the codewords do not introduce ambiguities at the BS in regards to which codewords were transmitted, i.e., based on the observation of the preambles in the virtual frame, the BS can always discern which codewords were actually sent. In the proposed scheme, such ambiguities exist - based on the observation, the BS may conclude that there are more codewords present in the virtual frame than there were actually sent. Namely, multiple combinations of transmitted codewords can yield the same observation, introducing phantom codewords which were not sent by any of the transmitting M-UEs. In Fig. 1(d) is depicted a combination of the three codewords used by M-UEs, which misleads the BS to perceive the phantom codeword \((B, A)\). The existence of phantom codewords affects adversely the performance, however, we show that this is significant only when the network is experiencing low user loads. Specifically, we show that the efficiency of the proposed approach (i.e., the fraction of M-UEs that successfully contended for access) for moderate and high loads and for the same number of used preambles and sub-frames in the virtual frame, is substantially higher than the efficiency of the reference scheme, despite the drawbacks caused by phantom codewords. We also show that by choosing the operating random access method according to the user load, it is possible to maintain an efficient random access over a large load region using the same number of preambles and sub-frames.

The remainder of this paper is organized as follows. In Section II we model and analyse both the reference and the proposed random access scheme. In Section III we discuss how the proposed code-expanded scheme can be used to enable a random access mechanism which adapts according to the load in the random access. Finally, Section IV concludes the paper.

II. SYSTEM MODEL

A. Preliminaries

In LTE the random access is performed through the Random Access Channel (RACH), which is mapped in the physical layer to the Physical Random Access Channel (PRACH) [10]. To start the asynchronous access each M-UE first selects one of the designated preambles and then waits until the next available PRACH sub-frame. The M-UE then performs the random access procedure, as described in [11], [12]. The periodicity of the PRACH resources can scale according to the expected RACH load. Therefore, the PRACH resources can occur from once in every sub-frame to once in every two frames, where a frame consists of 10 sequential sub-frames [13].

As outlined in Section II we assume that a group of consecutive sub-frames is organized in a virtual frame reserved for PRACH (i.e., random access), as depicted in the Fig. 2. We denote the number of sub-frames within a virtual frame by \(L\) and the number of available preambles by \(M\).

B. Reference Random Access Scheme

The reference scheme considered here is inspired by the LTE random access [11], [12]. We model it as an access reservation scheme [14], as depicted in Figure 3. We focus on the contention phase, which consists of \(L\) sub-frames (contention slots), and these sub-frames constitute a single virtual frame. We assume that there are \(N\) contending M-UEs; each M-UE randomly choses a preamble and a sub-frame of

---

**Fig. 2.** Example of PRACH opportunities organized in virtual frames.
the virtual frame in which the chosen preamble is sent.

The fraction of the M-UEs that succeed in the contention phase is denoted as the efficiency of the contention phase $S_r(N)$, and it is given by:

$$S_r(N) = \frac{N_S}{N_S + N_C}, \quad (1)$$

where $N_S$ is the number of codewords used by a single M-UE (singles) and $N_C$ is the number of codewords used by multiple M-UEs (collisions). From now on, we assess $N_S$ and $N_C$ by their expected values.

Assuming that the number of M-UEs contending per slot is modeled by the random variable $X$, then the probability that in a given contention slot there are $k$ M-UEs contending is:

$$Pr[X = k] = \binom{N}{k} \left( \frac{1}{A_r} \right)^k \left( 1 - \frac{1}{A_r} \right)^{N-k}, \quad (2)$$

where $A_r$ is the number of codewords available in the reference scheme. As outlined earlier, in a contention phase consisting of $L$ sub-frames, each M-UE sends only one preamble, therefore the number of available codewords is given by:

$$A_r = M \cdot L. \quad (3)$$

The expected number of codewords chosen by a single M-UE, $N_S$, is given by:

$$N_S = Pr[X = 1] \cdot A_r = N \left( 1 - \frac{1}{A_r} \right)^{N-1}, \quad (4)$$

and, similarly, the expected number of codewords chosen by multiple M-UEs, $N_C$, is:

$$N_C = Pr[X > 1] \cdot A_r = \left[ 1 - Pr[X = 0] - Pr[X = 1] \right] \cdot A_r = \left[ 1 - \left( 1 - \frac{1}{A_r} \right)^N - \frac{N}{A_r} \left( 1 - \frac{1}{A_r} \right)^{N-1} \right] A_r. \quad (5)$$

Substituting (4) and (5) in (1), we get the estimate of the efficiency of the contention phase.

### C. Code-Expanded Random Access

In the code-expanded random access, the outcome of the contention phase is the set of the codewords perceived by the BS, and all the perceived codewords are assumed to belong to the M-UEs that successfully contended for the access. However, the set of perceived codewords actually contains the codewords that are used just by a single M-UE (singles), codewords that are used by multiple M-UEs (collisions) and codewords that are used by none of the M-UEs (phantom codewords, see Fig. 1). The efficiency of the code-expanded random access $S_e(N)$ is:

$$S_e(N) = \frac{N_S}{N_P}, \quad (6)$$

where the $N_S$ is the number of single codewords, and $N_P$ is the expected number of codewords that the BS perceives, i.e., $N_P$ accounts for all single, collision and phantom codewords, where $N_P \leq A_e$.

The distribution of devices contending per virtual frame is modeled in the same way as in (2), with the difference that the number of available codewords is now higher. In the code-expanded approach, the M-UEs send in each sub-frame of the virtual frame either one of the $M$ preambles or the idle preamble $I$, so the total number of available codewords is:

$$A_e = (M + 1)^L - 1, \quad (7)$$

where the all-idle codeword is excluded.

The expected number of codewords chosen by a single M-UE, $N_S$, is given by:

$$N_S = Pr[X = 1] \cdot A_e = N \left( 1 - \frac{1}{A_e} \right)^{N-1}. \quad (8)$$

The method to obtain $N_P$ is discussed in the next subsection.

### D. Calculation of $N_P$

For the calculation of $N_P$ we use a representation based on a Markov Chain (MC) that describes the evolution of the perceived number of codewords by the BS when the number of contending M-UEs increases sequentially from 1 to $N$. We note that it is assumed that the M-UEs select their codewords independently and uniformly at random from the set of available codewords. The MC states are determined by the configuration that corresponds to the number of the observed preambles in the sub-frames of the virtual frame, which is created by the actual codewords selected by the M-UEs. The configuration is denoted by $(C_1, C_2, \ldots, C_L)$, where $C_j$ is the number of observed preambles by the BS in the $j$-th sub-frame. We note here that $C_j$ always includes the idle preamble, i.e., if the number of unique preambles in the $j$-th sub-frame is $C_j$, then the number of actually observed preambles is $C_j - 1$. Each state is characterized by a cardinality, which is the cardinality of the set of codewords perceived by the BS that is created by the given configuration. For $i$-th state configuration $(C_1, C_2, C_3, \ldots, C_L)$ the corresponding cardinality is:

$$\alpha_i = \prod_{j=1}^L C_j - 1, \quad (9)$$

where, once again, we assumed that all-idle codeword is not used by any of the M-UEs. Using this model, $N_P$ can be assessed as the average cardinality of the set of the codewords perceived after $N - 1$ transitions of MC, averaged over the probability distribution of the MC states after $N - 1$ transitions.

To ease the explanation, we focus on an example case where $L = 2$ and $M = 2$, and therefore $A_e = 8$. The full codebook
is shown in Table I while the MC representation including the state configurations, cardinalities and possible state transitions is shown in Table II. For example, the state configuration of the state 2 is \((1,3)\), implying that in the first sub-frame of the virtual frame there is an idle preamble \(I\), and in the second sub-frame there is an idle preamble and both available preambles, \(A\) and \(B\). The cardinality of state 2 is then \(\alpha_2 = 2\).

Initially, when there is only one M-UE attempting random access, the system can be in the states 1, 3 and 4. The probability of the system being in any of those states is simply the ratio of the number of codewords from the codebook which provide the corresponding state configuration and the total number of available codewords \(A_e\). For example, for state 1, where the state configuration is \((1,2)\), from Table I it can be seen that there are only two codewords that satisfy the state configuration, which are \((I,A)\) and \((I,B)\). Therefore, the probability of the BS perceiving this state upon the M-UE transmission of one of the available codewords is \(\frac{2}{8} = \frac{1}{4}\). A similar reasoning is done for the remaining possible initial states, and the following initial state probability vector \(\pi^{(1)}\) is:

\[
\pi^{(1)} = \begin{bmatrix} 1/4 & 0 & 1/4 & 2/4 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{10}
\]

where \(\pi^{(1)}_i\) is the probability that the system is initially in \(i\)-th state. Therefore, when one M-UE attempts random access, i.e. when \(N = 1\), \(N_P\) is obtained by:

\[
N_P = \sum_{i=1}^{A_e} \alpha_i \cdot \pi^{(1)}_i = 2 \tag{11}
\]

where \(\alpha_i\) is the cardinality of the \(i\)-th state, obtained from (9) and listed in Table II.

In case where there are two M-UEs attempting random access, it is assumed that the selection of the transmitted codewords is sequential and independent. Therefore, the codeword selected by the first M-UE leads the system to be in one of the possible initial states, which are 1, 3 and 4. When the second M-UE selects the codeword, the system can transit to any of the states listed in Table II.

The transition probabilities can be obtained following the same reasoning as the one used to obtain \(\pi^{(1)}\). Consider that the first M-UE selects a codeword that puts the system in state 1, i.e. the M-UE selected either \((I,A)\) or \((I,B)\). Now, whatever codeword the second M-UE selects, the system can only transit to states 1, 2, 4, 5. For the system to transit from state 1 to state 5, it means that the second M-UE has to select a codeword that consists of either preamble \(A\) or \(B\) in the first sub-frame, and the remaining yet unused preamble in the second sub-frame (i.e. the codeword \((A,B)\) or \((B,A)\)) if the initial codeword was \((I,A)\), or the codeword \((A,A)\) or \((B,B)\) if the initial codeword was \((I,B)\), thus making the configuration become \((2,3)\). From the preceding discussion, we see that no matter which codeword caused the chain to be in state 1, the transition to state 5 can be caused by the second M-UE selecting (one of) two codewords from the set of all codewords. Therefore, this transition probability is equal to \(\frac{2}{8} = \frac{1}{4}\) and it does not depend on which codeword the first M-UE selected to bring the system to state 1. Using similar reasoning, it can be shown that the transition probabilities do not depend on the choices of codewords that brought the system to a given state and, for every state transition, the transition probability is the ratio of the number of codewords that enable the transition and the total number of available codewords \(A_e\). For the considered example, the transition matrix \(P\) is:

\[
P = \frac{1}{8} \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}
\]

where \(P_{ij}\) is the transition probability from \(i\)-th to \(j\)-th state. The expected number of perceived codewords \(N_P\) can be obtained as follows:

\[
N_P = \sum_{i=1}^{A_e} \alpha_i \cdot \pi^{(N)}_i, \tag{12}
\]

where \(\pi^{(N)}_i\) is the \(i\)-th element of the state vector \(\pi^{(N)}\), i.e., the probability of the BS perceiving a state configuration given by the \(i\)-th state after the \(N\)-th user has chosen his codeword. The \(\alpha_i\) is the cardinality of the \(i\)-th state, given by (9). As for any MC, the state probability distribution \(\pi^{(N)}\) is:

\[
\pi^{(N)} = \pi^{(1)} \cdot P^{(N-1)}. \tag{13}
\]

Fig. 4 depicts the comparison of the efficiencies for the reference scheme calculated using (1). code-expanded scheme.
calculated using (6) and (12), and code-expanded scheme that is obtained using Monte Carlo simulations, for the example case when \( L = 2 \) and \( M = 2 \). It is obvious that, as the number of M-UEs grows, the efficiency of the code-expanded method outperforms the reference one. Also, both curves corresponding to the code-expanded method converge, justifying the use of (12) when computing the efficiency.

Finally, we note that the described methodology for modeling \( N_P \) can be generalized for arbitrary \( L \) and \( M \).

III. ADAPTIVE CODE-EXPANDED RANDOM ACCESS

In the previous subsection we considered using only the codebook consisting of all the available codewords (i.e., the full codebook). In this subsection we consider the case when a subset of the full codebook is used, where a subset is obtained by restricting the number of preambles that could be used in each of the sub-frames of the virtual frame (we note that these restrictions are not necessarily the same for every sub-frame). The efficiency in this case can be computed in the same way as in the previous subsection; we omit the corresponding analysis due to space constraints.

As an illustrative example consider the case where \( L = 2 \) and \( M = 4 \), for which the MC model is given in Table III. The possible cardinality values of the states for this MC are then \{1, 2, 3, 4, 5, 7, 8, 9, 11, 14, 15, 19, 24\}. This sequence represents all the possible subsets in regards to the number of codes that can be used. For \( L = 2 \) and \( M = 4 \) the number of codewords for the reference scheme is \( A_r = 8 \), so the cardinalities of interest are the ones where \( A_e = \{9, 11, 14, 15, 19, 24\} \). In Fig. 5 is depicted the efficiency for the given example, both for the reference scheme and for the code-expanded scheme when cardinalities of interest are used. As it can be observed, there is a set of threshold values in regards to the number of M-UEs, where the system should increase the cardinality of the codebook in use, in order to maintain the efficiency. Overall, it could be concluded that an adaptive approach, where the codebook is selected based on the estimate of the user load is the preferred operating strategy.

In Fig. 5 is depicted the efficiency of the adaptive code-expanded approach for \( L = 4 \) and \( M = 4 \). Comparing these results with the results given in Fig. 5 it can be observed that the region where the gain of the adaptive approach over the case when the full codebook is used is more obvious. Also, the increased \( M \) provides for significant expansion of the supported user load region.

Another way to exploit the code-expanded scheme is to use a significantly lower number of preambles per sub-frame than in the reference scheme, while reaching the same efficiency, since the preamble detection performance decreases with the number of simultaneous preambles in a sub-frame [15]. In the code-expanded scheme the number of available codewords scales exponentially with the virtual frame length [7] while in the reference schemes it scales linearly [5]. Assuming that the number of preambles in use in the reference and code-expanded schemes are denoted as \( M_r \) and \( M_e \), respectively, the lower bound on \( M_e \) required to outperform the reference random access occurs when \( A_e > A_r \) and can be calculated.
using (7) and (3) as:

\[ M_e > \left\lfloor \sqrt{Mr \cdot L + 1} - 1 \right\rfloor \].

As example consider the plots in Figures 6 and 7 where \( L = 4 \), \( Mr = 32 \) and \( Me \in \{3, 4\} \) \( (Me = 3 \) is the lower bound obtained by (14)). In case when \( Me = 3 \) the code-expanded random access achieves higher performance than the reference scheme when number of users is \( N > 225 \), and the supported user load region is increased several times. If the number of available preambles is increased just by one, i.e., \( Me = 4 \), the threshold value for \( N \) stays the same (which is in line with results presented in Fig. 5), but the supported load region is extended significantly further.

The above results suggest the approach in which subsets of available preambles are reserved for different user classes, thus allowing user separation. The number of preambles per class should be selected according to expected user loads and traffic types, such that a satisfactory efficiency is achieved.

**IV. Conclusion**

In this paper we proposed a code-expanded random access inspired by the LTE random access. The proposed scheme increases the amount of available contention resources, without resorting to the increase of system resources, such as contention sub-frames and preambles. This increase is accomplished by expanding the contention space to the code domain, through the creation of random access codewords.

It was shown that for high user loads the proposed scheme is significantly more efficient than the reference scheme. Also, it was shown that by selecting the appropriate number of random access codewords it is possible to maintain the random access scheme efficiency over a large load region. This suggests the usage of an adaptive random access scheme, i.e., a combination of the reference and code-expanded random access, which allows to maintain the efficiency both for low and high user load regions.

The proposed random access scheme can be further complemented by creating user classes, which are separated by the selection of used preambles. Specifically, due to amount of contention codewords that can be obtained with a low number of preambles and that could be reserved for MTC users, the remaining preambles could be used by the human-centric users, thus allowing human-centric and device-centric traffic to coexist.

Finally, we note that for optimal operation of the proposed expansions, it is necessary to obtain the estimates of the user loads. The design of the adaptive random access scheme that implements the estimation of the user load is a topic of ongoing research.

**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] R. Hu, Y. Qian, H.-H. Chen, and A. Jamalipour, “Recent progress in machine-to-machine communications [guest editorial],” Communications Magazine, IEEE, vol. 49, pp. 24–26, april 2011.

[2] M. J. Anthony Lo, Yee Wei Law and M. Kucharczak, “Enhanced In-advanced random-access mechanism for massive machine-to-machine (m2m) communications,” in 27th World Wireless Research Forum (WWRF) Meeting, 2011.

[3] 3GPP TR 37.868 V11.0, Study on RAN Improvements for Machine-type Communications, October 2011.

[4] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, “Overload control for machine-type-communications in lte-advanced system,” Communications Magazine, IEEE, vol. 50, pp. 38 –45, June 2012.

[5] A. X. L. J. P. M. Zubair Shafiq, Lusheng Ji and J. Wang, “A first look at cellular machine-to-machine traffic - large scale measurement and characterization,” in Proceedings of the International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), (London, United Kingdom), June 2012.

[6] 3GPP TS 22.368, Service requirements for MTC; Stage 1.

[7] S.-Y. Lien, T.-H. Liu, C.-Y. Kao, and K.-C. Chen, “Cooperative access class barring for machine-to-machine communications,” Wireless Communications, IEEE Transactions on, vol. 11, january 2012.

[8] J.-P. Cheng, C. han Lee, and T.-M. Lin, “Prioritized random access with dynamic access barring for ran overload in 3gpp lte-a networks,” in GLOBECOM Workshops, 2011 IEEE, pp. 368 –372, dec. 2011.

[9] K.-D. Lee, S. Kim, and B. Yi, “Throughput comparison of random access methods for m2m service over lte networks,” in GLOBECOM Workshops, 2011 IEEE, pp. 373 –377, dec. 2011.

[10] 3GPP TS 36.212, Multiplexing and channel coding.

[11] 3GPP TS 36.321, Medium Access Control (MAC) protocol specification.

[12] 3GPP TS 36.213, Physical layer procedures.

[13] 3GPP TS 36.211, Physical Channels and Modulation.

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

[15] Nortel Networks R1-062139, On the signal structures of non-synchronized RACH, 2006.