Cluster-based Group Paging Scheme with Preamble Reuse for mMTC in 5G Networks

Jiajie Huang, Wen Zhan, Xinghua Sun, Pei Liu, and Dejin Kong

Abstract—The massive Machine Type Communications (mMTC) is one of the three generic services to be supported by 5G wireless systems. To fulfill the ever-increasing network access demand from a large number of Machine Type Devices (MTDs), this paper develops a cluster-based group paging scheme. Specifically, with the proposed scheme, MTDs are divided into clusters and the group paging period is decomposed into two parts: intra-cluster access period and inter-cluster access period. In the intra-cluster access period, preamble reuse is adopted for facilitating the access request transmissions from MTDs in each cluster to its cluster head. In the inter-cluster access period, only cluster heads send access requests to the base station.

To evaluate and optimize the access efficiency of the proposed scheme, the probability of successful access of each MTD is characterized, based on which the maximum probability of successful access and the corresponding optimal number of clusters and optimal length of the intra-cluster access period are obtained as explicit functions of key system parameters including the number of preambles and the number of MTDs. A comparative study of the access efficiency for group paging with clustering and without clustering is conducted, which reveals the critical threshold in terms of the number of MTDs, above which clustering is beneficial. The analysis is verified via extensive simulations. It is shown that the access performance of the proposed scheme significantly outperforms that of the traditional group paging scheme, especially in massive access scenarios.

Index Terms—5G, group paging, cluster, massive Machine Type Communications (mMTC).

I. INTRODUCTION

Machine Type Communications (MTC), also known as Machine to Machine (M2M) communications, is a new type of communication paradigm in which devices can automatically operate, communicate, process information without or with few human intervention. It is the key technology for enabling many emerging Internet of Thing (IoT) applications such as smart city, smart agriculture and e-Health. Recent reports have revealed that the number of Machine Type Devices (MTDs) will surpass 10 billion in the next few years and the data traffic from MTDs will constitute the major portion of the traffic volume in the next-generation cellular networks [1]. Therefore, the cellular standardization body, the 3rd Generation Partnership Project (3GPP), has identified mMTC as the one of the three generic services to be supported by 5G system [2].

Supporting mMTC in 5G networks has to address many challenges, one of which that captures much attention is the severe congestion issue in the cellular random access channel [3]. Specifically, in cellular networks, each MTD has to establish a connection with the BS via the random access procedure, before its data transmission. In the random access procedure [4], each MTD randomly selects one orthogonal preamble from the preamble resource pool and transmits via the shared random access channel. If more than one MTDs transmit the same preamble at the same time, a collision occurs, in which case all access requests fail. Due to limited number of orthogonal preambles in cellular systems, when massive MTDs transmit access requests, frequent preamble collisions occur and most of access requests fail, leading to a very low access efficiency.

To address the above issue, group paging has been identified by 3GPP as one standardized solution [5]. With group paging, the BS would put MTDs into different groups according to various metrics such as Quality-of-Service (QoS) requirement or application type, and labels each MTD in the same group with a certain group identifier. If the BS needs to collect information from a group, then it sends a paging message enclosed with the corresponding group identifier, and indicates a period of time, which is referred to as the paging period. MTDs in this group will then attempt to access the BS and send messages in this specified period. As the BS determines when MTDs could access the network, group paging enables flexible control of MTC traffic and particularly suits for duty-cycle MTC applications. Yet, various analytical results in the existing literature have found that if the group size is large, then the access efficiency with group paging is also intolerably low [6], [7]. The fundamental reason, nevertheless, lies again in the massive access requests from MTDs during the paging period.

To reduce the number of concurrent access requests during the paging period, cluster-based random access scheme could be a promising way. The basic idea is further dividing MTDs into clusters, where in each cluster, one MTD is selected as the Cluster Head (CH) and others are Cluster Members (CMs). CMs send access requests to the CH, and the CH would relay the access requests to the BS. It is clear that in each cluster, only the delegated MTD, i.e., CH, sends the access request to
the BS. Therefore, the number of concurrent access requests could be significantly reduced.

There have been lots of works on cluster-based random access for mMTC in cellular networks. For instance, in [8] and [9], MTDs are divided into two clusters depending on different QoS requirements, and preambles are adaptively allocated for the cluster with stringent QoS requirement. In [10], only delay-tolerant MTDs are clustered, while delay-sensitive MTDs send access requests directly to the BS. In [11], an analytical model is developed to evaluate the reliability of device-to-device links in each cluster and its effect on the access performance. In [12]–[15], data aggregate process is considered, where cluster members transmit packets to the cluster head first, and then the cluster head initiates the access request transmission. In [16] and [17], it is revealed that when clusters are dispersed widely, preamble reuse among clusters is possible, which can further improve the overall access performance. In a nutshell, existing results have validated that by adopting the clustering technique, the access efficiency for mMTC in cellular networks can be significantly boosted.

Regarding this, we will enhance the access performance of MTDs in group paging by using the clustering technique, where a novel cluster-based group paging scheme is proposed in this paper. The basic idea is to partition the group paging period into two parts: intra-cluster access period and inter-cluster access period. In the intra-cluster access period, CMs send access request to a delegated CH; in the inter-cluster access period, CHs send access requests to the BS. To facilitate the intra-cluster access, preamble reuse is adopted, for which we introduce a preamble reuse factor to characterize the preamble reuse level in network-wide.

To evaluate the access efficiency of the proposed scheme, the probability of successful access of each MTD is derived as an explicit function of key system parameters including the preamble reuse factor, the number of clusters, the length of group paging period and the length of intra-cluster access period. To exploit the limiting performance of the proposed cluster-based group paging scheme, the maximum probability of successful access is obtained by jointly tuning the number of clusters and the length of intra-cluster access period. The analysis sheds important light on practical system design. It is revealed that the access performance of group paging with clustering outperforms that without clustering only if the number of MTDs in the group is larger than a certain threshold. Such a critical threshold is explicitly characterized in this paper and shown to be solely determined by the preamble reuse factor.

The rest of the paper is organized as follows. Section II introduces the network scenario. A novel cluster-based group paging scheme with preamble reuse is proposed in Section III. In Section IV, we derive the maximum probability of successful access and corresponding optimal system parameter setting. In Section V, simulation results are presented to validate the analysis. Concluding remarks are summarized in Section VI.

II. SYSTEM MODEL AND PRELIMINARY ANALYSIS

This study considers a cellular network with one BS and one paging group. The group size, i.e., the total number of MTDs in this group, is \( N \). We consider a general scenario, in which the BS sends a paging message to this group and identifies a paging period with length \( T \) time slots. Each time slot denotes a time period in which one time-frequency uplink random-access resource unit is allocated, where MTDs can transmit preambles. Upon the reception of paging message, all MTDs in the group would perform the random access procedure during the paging period.

According to the standards [4], in the random access procedure, each MTD randomly selects one out of \( M > 1 \) orthogonal preambles and transmits it to the BS. The access request transmission is successful if and only if there is no concurrent transmission of a given preamble at the same time slot. Otherwise, a collision happens and all of them fail. In this paper, we consider the one-shot transmission of each request, where each MTD just transmits the access request only once. We also consider the pre-backoff scheme for each MTD, that is, each MTD randomly chooses one among \( T \) time slots for transmitting the access request\(^1\).

Let us first consider the case that each MTD transmits access request to the BS directly. In this case, the probability of successful access of each MTD in group paging is given by

\[
p_{\text{direct}} = \left(1 - \frac{1}{MT} \right)^{N-1}. \tag{1}
\]

We can clearly see from (1) that when the group size \( N \) is large, the probability of successful access would be very small.

\(^1\)It has long been observed that upon the reception of paging message, if all MTDs send access requests at the first time slot of the paging period, then the access efficiency would be very low due to frequent preamble collision. With pre-backoff scheme, bursty access requests could be dispersed over the time domain to reduce the number of concurrent requests, which may improve the access efficiency [18].
III. CLUSTER-BASED GROUP PAGING SCHEME WITH PREAMBLE REUSE

To improve the access performance, in this section, we will incorporate the clustering method into the group paging, and propose a cluster-based group paging scheme. As Fig. 1 illustrates, MTDs are divided into \( C \) clusters. In each cluster, one MTD with the capability of caching access requests is selected as the CH, and other MTDs are referred to as CMs. The group paging period is divided into two parts: intra-cluster access period and inter-cluster access period, as Fig. 2 illustrates, where the length of intra-cluster access period is \( t \) time slots, \( t \in \{1, 2, \ldots, T - 1\} \). In the intra-cluster access period, CMs in each cluster send access requests to their CH; In the inter-cluster access period, CHs send access requests to the BS. It is clear that the access request of an MTD can be successfully received by the BS if and only if both the intra-cluster access and the inter-cluster access are successful, i.e., its CH successfully receives its access request and the BS successfully receives its CH’s access request.

Note that the system would assign a set of preambles for each cluster for the use of intra-cluster access. Reusing the same set of preambles in different clusters is practically feasible, because those clusters may be distributed apart from each other such that the interference during preamble transmission becomes negligible [16], [17]. Without loss in generality, in this paper, we introduce a preamble reuse factor \( \mu \) to characterize the preamble reuse level in network-wide, where the number of available preambles for each cluster is given by \( M \cdot \mu \). We assume \( \mu \) is a given system parameter, where \( \mu \in \left[ \frac{1}{C}, 1 \right] \), and \( \mu = \frac{1}{C} \) represents that preambles are not reused, i.e., each cluster uses different set of preambles and each set contains \( \frac{M}{C} \) preambles, \( \mu = 1 \) represents that preambles are fully reused, i.e., each cluster can use all \( M \) preambles. An example will be presented in Section V to show how the reuse factor is determined.

When MTDs are uniformly distributed and the coverage of each cluster is similar, we approximate the number of MTDs in each cluster as \( \frac{N}{C} \). The probability that each CM successfully sends an access request to the CH can then be obtained as

\[
   p_{\text{intra}} = \left( 1 - \frac{1}{M \cdot \mu \cdot t} \right)^{\frac{N}{C}}. \tag{2}
\]

As the inter-cluster access period lasts for \( T - t \) time slots, the probability that each CH successfully sends an access request to the BS is given by

\[
   p_{\text{inter}} = \left( 1 - \frac{1}{M \cdot (T - t)} \right)^{\frac{C}{t}}. \tag{3}
\]

Let \( p_{\text{cluster}} \) denote the probability of success access of each MTD in the proposed cluster-based group paging scheme. Since the access request of each MTD can be successfully received by the BS if and only if both the intra-cluster access and the inter-cluster access are successful, by combining (2)–(3), we can then have

\[
   p_{\text{cluster}} = p_{\text{intra}} \cdot p_{\text{inter}} = \left( 1 - \frac{1}{M \cdot \mu \cdot t} \right)^{\frac{N}{C}} \cdot \left( 1 - \frac{1}{M \cdot (T - t)} \right)^{\frac{C}{t}}. \tag{4}
\]

Fig. 3a demonstrates how the probability of successful access \( p_{\text{cluster}} \) varies with the number of clusters \( C \) with the length of intra-cluster period \( t = 4, 8, 12, 16 \), and the number of MTDs \( N = 10000 \). It can be seen from Fig. 3a that for given \( t \), as the number of clusters \( C \) increases, the probability of successful access \( p_{\text{cluster}} \) increases first and then drops. Intuitively, when \( C \) is small, each cluster contains a large number of MTDs, leading to an intensified inter-cluster contention and therefore a small \( p_{\text{intra}} \). On the other hand, when \( C \) is large, the inter-cluster contention becomes fierce, which results in a small \( p_{\text{inter}} \). Accordingly, for given \( t \), the number of clusters \( C \) should be carefully selected for maximizing the probability of successful access \( p_{\text{cluster}} \). Similar observation can also be obtained from Fig. 3b, which shows how \( p_{\text{cluster}} \) varies with the length of intra-cluster period.
period \( t \). From Fig. 3a-b, we can conclude that the probability of successful access \( p_{\text{cluster}} \) is crucially determined by \( C \) and \( t \). Therefore, in the following section, we study how to tune \( C \) and \( t \) to maximize \( p_{\text{cluster}} \).

IV. PERFORMANCE OPTIMIZATION

In this section, we aim to optimize the performance of the proposed cluster-based group paging scheme. Specifically, we are interested in maximizing the probability of successful access of each MTD \( p_{\text{cluster}} \) by properly choosing the length of the intra-cluster access period \( t \) and the number of clusters \( C \) given the length of the paging period \( T \), the preamble reuse factor \( \mu \), the number of preambles \( M \) and the number of MTDs \( N \). The corresponding optimization problem can be written as

\[
p_{\text{opt}} = \max_{\{t,C\}} \left( 1 - \frac{1}{M \mu^t} \right)^N \left( 1 - \frac{1}{M(T-t)} \right)^{C-1},
\]

s.t. \( 1 \leq t \leq T - 1 \),
\( 1 < C \).

Let \( p_{\text{opt}} \) denote the maximum probability of successful access and, \( t_{\text{opt}} \) and \( C_{\text{opt}} \) denote the corresponding optimal length of intra-cluster access period and optimal number of clusters, respectively. The following theorem presents \( p_{\text{opt}} \), \( t_{\text{opt}} \) and \( C_{\text{opt}} \).

**Theorem 1.** The maximum probability of successful access is given by

\[
p_{\text{opt}} = \exp \left( -\frac{1}{MT} \sqrt{\frac{N}{\mu}} \right),
\]

which is achieved when the length of intra-cluster access period is set to

\[
t_{\text{opt}} = \frac{T}{2},
\]

and the number of clusters is set to

\[
C_{\text{opt}} = \sqrt{\frac{N}{\mu}}.
\]

**Proof.** See Appendix A.

Theorem 1 shows that the maximum probability of successful access \( p_{\text{opt}} \) increases with the number of preambles \( M \), the length of paging period \( T \) and the preamble reuse factor \( \mu \). It is intuitively clear that with more resources in preamble domain or the time domain, the access performance should be improved. At the meantime, it is surprisingly to see from Theorem 1 that the optimal length of intra-cluster access period \( t_{\text{opt}} \) is always half of \( T \), which has no concern with \( M, \mu \) and the number of MTDs \( N \). The reason leads to (7) comes from (8), where the number of clusters \( C \) is optimally tuned according to \( \mu \) and \( N \), such that in both intra-cluster access period and inter-cluster access period, the ratio of the number of MTDs over the amount of resource in preamble domain reaches the same level. With \( t = t_{\text{opt}} \) and \( C = C_{\text{opt}} \), we can obtain that

\[
p_{\text{intra}} = p_{\text{inter}} = \exp \left( -\frac{\sqrt{N/\mu}}{M} \frac{T}{2} \right),
\]

with which the maximum probability of successful access \( p_{\text{opt}} \) is achieved.

To see the performance gain brought by clustering, Fig. 4 shows how the ratio \( p_{\text{opt}} / p_{\text{direct}} \) varies with the number of MTDs \( N \) with the preamble reuse factor \( \mu \in \{0.2, 0.4, 0.6, 0.8\} \), where \( p_{\text{direct}} \) in (1) denotes the probability of successful access without clustering. It can be seen from Fig. 4 that for given \( N \), \( p_{\text{opt}} / p_{\text{direct}} \) increases with \( \mu \). Obviously, a larger preamble reuse factor \( \mu \) leads to more available preambles for each cluster, and improves both the maximum probability of successful access with clustering \( p_{\text{opt}} \) and the ratio \( p_{\text{opt}} / p_{\text{direct}} \).

On the other hand, for given preamble reuse factor \( \mu \), \( p_{\text{opt}} / p_{\text{direct}} \) decreases slightly first and then increases as the group size \( N \) grows. Intuitively, when the group size is small, the channel contention is already light. Clustering introduces one more round of access request contention at the inter-cluster level, which, however, results in opposite effect. As the group size grows, the ratio \( p_{\text{opt}} / p_{\text{direct}} \) sharply increases, indicating that significant performance gain can be achieved by clustering. Let \( N_{\text{thr}} \) denote the threshold in terms of the number of MTDs, above which \( p_{\text{opt}} > p_{\text{direct}} \). According to (1) and (6), we can have

\[
N_{\text{thr}} = \frac{16}{\mu}.
\]

(10) shows that with a large preamble reuse factor \( \mu \), the threshold \( N_{\text{thr}} \) would be low, which implies that even when the group size is small, the access performance of group paging can still be improved via clustering.

V. SIMULATION RESULTS

In this section, simulation results are presented to validate the proceeding analysis. In simulations, we consider a single-cell scenario with the number of MTDs \( N \in [10^3, 10^4] \). The BS adopts the clustering algorithm that is based on geographical location information [19], which divides the area into regular hexagons. The MTDs in the same hexagon belongs to the same cluster [20], and one MTD near the center.
of the hexagon is selected as the CH. MTDs are uniformly distributed and therefore when the number of MTDs is large, the number of MTDs in each hexagon would be approximately same. The BS can tune the size of each hexagon so that the number of clusters, i.e., hexagons, for covering the given area can be adjusted. Each cluster is assigned a set of preambles, which contains completely different preambles than those for neighboring clusters. The transmission power of each MTD in each cluster is perfectly tuned so that there is no interference between the clusters which are not neighboring. Fig. 5 shows an example of the network scenario, in which three neighboring clusters collectively use the complete set of available preambles. The reuse factor \( \mu \) indicates, the system can enlarge the number of preambles \( M \), the length of paging period \( T \) or the preamble reuse factor \( \mu \). Simulation results presented in Fig. 7 confirm that under different values of \( N \), \( p_{\text{opt}} \) can be effectively improved with the increase of \( M, T \) or \( \mu \). Finally, simulation results presented in Figs. 6–7 well agree with the analysis.

**VI. CONCLUSION**

In this paper, we propose a cluster-based group paging scheme with preamble reuse for mMTC in 5G networks. The access performance of the network with the proposed scheme is evaluated based on the probability of successful access for each MTD. The explicit expression of this performance metric is derived, with which the maximum probability of successful access and corresponding optimal number of clusters and length of the intra-cluster access period are further characterized. The analysis is validated via extensive simulations. It is shown that with the proposed scheme, the access performance significantly outperforms that with the traditional group paging scheme when the number of MTDs is large, indicating that the proposed scheme is promising for being used in massive access scenarios. It is also shown that to further improve the maximum probability of successful access under the proposed scheme, the network could extend the length of the paging period, allocate more preambles or enhance the preamble reuse level in the network.

Note that in this paper, we evaluate the network performance by assuming that the number of MTDs in each cluster is approximately same. Yet, in practical case, when MTDs are not uniformly distributed, the number of MTDs in each cluster may quite different from each other. In future work, we will extend the analysis to incorporate such a heterogeneous scenario.

**APPENDIX A**

**PROOF OF THEOREM 1**

*Proof.* To derive the maximum probability of successful access and the corresponding optimal system parameter setting, let us first rewrite (4) as

\[
p_{\text{cluster}} = \left( 1 - \frac{1}{M \mu t} \right)^{N - 1} \cdot \left( 1 - \frac{1}{M \mu (T - t)} \right)^{C - 1} \approx \exp \left\{ - \left( \frac{N}{M \mu C T} + \frac{C}{M \mu (T - t)} \right) \right\}
\]

by applying the approximation \( y - 1 \approx y \) and \( (1 - x)^y \approx \exp (-xy) \) for large \( y \) and \( 0 < x < 1 \).
By letting \( \frac{\partial p_{\text{cluster}}}{\partial t} = 0 \), we have

\[
(\mu C^2 - N) \cdot t^2 + 2NT \cdot t - NT^2 = 0,
\]

where (12) is one variable quadratic equation about \( t \), and the solution is given by

\[
t = \frac{-2NT \pm 2T \sqrt{N/C}}{2(\mu C^2 - N)} = \frac{N - \sqrt{N/C}}{N - \mu C^2}.
\]

Since \( t = T \cdot \frac{N - \sqrt{N/C}}{N - \mu C^2} \) does not satisfy the constraint of \( 1 < t < T - 1 \), this solution is discarded. Therefore, we have the single feasible solution of (12) as

\[
t = T \cdot \frac{N - \sqrt{N/C}}{N - \mu C^2}.
\]

By substituting \( t \) into (11), we can rewrite the expression of \( p_{\text{cluster}} \) as

\[
p_{\text{cluster}} = \exp \left\{ \left( \frac{N(N - \mu C^2)}{MCT(N - \sqrt{N/C})} + \frac{C(N - \mu C^2)}{MT(N - \sqrt{N/C})} \right) \right\},
\]

and

\[
\frac{\partial p_{\text{cluster}}}{\partial C} = \frac{1}{M T} \left\{ \frac{2N \mu C}{\mu C(N - C \sqrt{N/C})} + \frac{2C}{\sqrt{N/C}} \right\}
\]

\[
+ \left( \mu C^2 - N \right) \left\{ \frac{2C}{\mu C(N - C \sqrt{N/C})} + \frac{\mu}{\sqrt{N/C}} \right\}.
\]

By solving \( \frac{\partial p_{\text{cluster}}}{\partial C} = 0 \), we can get (8). By substituting (8) in (14), we obtain (7). (6) can then be derived by combining (7), (8) and (11).

REFERENCES