SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Machine-Type Communications

October 2017, Vol. 60 100305:1–100305:10
doi: 10.1007/s11432-017-9196-0

Aggregation transmission scheme for machine type communications

Yanhuan SUN[1], Ming ZHAO[1*], Sihai ZHANG[1], Jinkang ZHU[2], Wuyang ZHOU[1] & Shengli ZHOU[3]

[1]*Key Laboratory of Wireless-Optical Communications, Chinese Academy of Sciences, University of Science and Technology of China, Hefei 230027, China;*
[2]*Laboratory of Personal Communication Network and Spread Spectrum, University of Science and Technology of China, Hefei 230027, China;*
[3]*Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA*

Received April 24, 2017; accepted August 2, 2017; published online September 1, 2017

Abstract Massive amount of small data generated by machine type communications (MTC) will pose a challenge to the future fifth generation (5G) wireless network. Since the information from or to the machine type users aggregating closely are highly correlated, the relevance of data can be excavated by big data analysis to help improve the spectral efficiency. In this paper we proposed an aggregation transmission scheme (ATS) for MTC downlink transmissions in which the transmission order of users' data packets can be adjusted according to their relevance under the delay constraints. The users having relevance will temporally share the time slots and their data are transmitted in a multicast way so that much less timeslots are needed. We propose three different algorithms, conditional random search (CRS), standard-row algorithm (SRA), and genetic algorithm (GA) to tackle the problem of transmission order adjustment. Simulation results validate the good performance of ATS and demonstrate that SRA has the lowest complexity while GA may achieve a better performance. We also analyze the impact of different delay requirements. Our work sheds light on dealing with massive MTC data traffic for future wireless communications.

Keywords machine type communications, aggregation transmission scheme, delay tolerance, genetic algorithm, big data

Citation Sun Y H, Zhao M, Zhang S H, et al. Aggregation transmission scheme for machine type communications. Sci China Inf Sci, 2017, 60(10): 100305, doi: 10.1007/s11432-017-9196-0

1 Introduction

The large amount of emerging intelligent devices are driving the wide range of potential applications and services such as healthcare, smart gird, tracking, environment monitoring, payment, vending machines, building automation, and etc. [1–4]. It is forecasted that about 50 billion machines will be connected to the networks in 2020 [5], mostly through wireless networks. So the large amount of machine type communications (MTC) traffic will pose a challenge to the future fifth generation (5G) systems [6–9].

* Corresponding author (email: zhaoming@ustc.edu.cn)

To meet the explosive resource demands in cellular systems, MTC is in great need of improving the utilization of the limited wireless resources.

Multiple candidate solutions have been investigated to fully utilize the wireless resources and some of them can be applied in MTC field. In [10] a new paradigm was proposed which was motivated by emerging systems with a massive number of users in an area and the capacity of gaussian many-access channels was analyzed theoretically. In [11], authors proposed a nonorthogonal many-access scheme that allows for a large number of users to transmit small packets simultaneously. In [12], cooperative relay is applied to make machine-type communications and cellular communications operate in orthogonal resource sharing way. In [13], a grouping-based radio resource management with considerately decreased computational complexity which can support an enormous number of MTC devices was proposed. Although all these schemes can improve the spectral efficiency greatly, they do not take the similarity of users' contents into account.

Multicast is a good scheme for improving the spectral efficiency. It is fully used by many operators to efficiently utilize the available bandwidth of their networks in delivering the same content to multiple receivers. The 3rd generation partnership project (3GPP) has already proposed one technology for long term evolution (LTE) called evolved multimedia broadcast and multicast services (MBMS) [14]. Because of its feasibility, MBMS can be applied to many fields such as traffic status reports [15], local news, weather forecasts, stock market, advertising, media and entertainments such as IPTV, mobile TV, video conferences, and etc. However, MBMS is only suitable when different users' contents remain exactly the same for a long time and it does not consider the delay tolerance of users.

We notice that an important behavior pattern of these users is that the densely aggregating users may request similar or correlated content almost at the same time in one cell. Fully using the similarity or correlation of users' request can save more resources further. Actually MTC devices' transmissions can endure certain latency, from seconds to minutes. If we can fully use their delay requirements, the transmission order can be adjusted. Moreover, big data will bring a lot of opportunities in 5G wireless communications [16]. Using properly designed prediction techniques can produce prediction accuracy over 90 percentage, like the model shown in [17, 18]. If we make full use of massive wireless big data gathered from telecommunication operators, internet service providers and other data sources, the correlation among MTC users' traffic and delay requirements can be learned, predicted and inferred by machine learning and prediction techniques.

Based on the above consideration, we proposed an aggregation transmission scheme (ATS) which adjust the transmission order of users' data packets under the delay constraints by exploiting the similarity and relevance of the data, so the data of all users can be transmitted in a temporally multicast way to save the radio resource. Compared with current transmission mode and traditional MBMS, the proposed ATS can achieve a much higher transmission efficiency.

Our contributions in this paper have two fold:

- We proposed a new ATS dealing with future massive connectivity in MTC field. ATS can help save the radio resources by exploiting the similarity and relevance of all users' data under the delay constraints.
- We proposed three algorithms, conditional random search (CRS), standard-row algorithm(SRA) and genetic algorithm(GA), for adjusting the data packets transmission order in ATS. We reveal the complexity and performance of these algorithms with simulations.

2 System model

Consider an MTC user cluster with M ($M \gg 1$) users gathering in a region covered by a wireless base station (BS). Assume the BS has to transmit $M \times K$ data blocks to M users in K time slots. The transmission order of each user follows identical distribution. Let B_m^k denote the data block required by user m ($m \leqslant M$) in the kth time slot. The BS is supposed to transmit the data to user m in the order as $I_m = (B_m^1, B_m^2, \ldots, B_m^K)$, and the transmission orders for all users can be denoted as a matrix

$$\boldsymbol{I} = \left(I_1^{\mathrm{T}}\ I_2^{\mathrm{T}}\ \ldots\ I_M^{\mathrm{T}} \right)^{\mathrm{T}} = \begin{pmatrix} B_1^1 & B_1^2 & \ldots & B_1^K \\ B_2^1 & B_2^2 & \ldots & B_2^K \\ \vdots & \vdots & \vdots & \ddots \\ B_M^1 & B_M^2 & \ldots & B_M^K \end{pmatrix}.$$

Since the users aggregating in a cluster may request some highly correlated information, they have some common data packets. Therefore, it will be a good approach to save the resource consumption by transmitting these data in a multicast way. That motivates us to come up with the ATS in this paper. However, these common data requested by different users may not be aligned in time. So in order to aggregate these common data, we need to evaluate the relevance of the data packets of different users and their transmission delay requirements.

2.1 Information relevance

Consider users i and j transmitting two correlated data flows over K timeslots, i.e. some data blocks of the two flows are the same. To quantitatively describe the similarity or correlation of users' information sequences, we define a new metric as information relevance (IR)

$$Q_{i,j} = \frac{\sum_{k=1}^{K} Q_{i,j,k}}{K}, \tag{1}$$

where $Q_{i,j,k}$ indicates the block-level similarity of user i and j on the kth timeslot and it equals to 1 for that the two blocks are identical and 0 for otherwise. We further define the IR of all M users as

$$Q = \frac{\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} Q_{i,j}}{(M \times (M-1))/2}. \tag{2}$$

The higher is the value of Q the more similarities all users share in data flows, which offers more possibility for ATS to save wireless resources.

2.2 Delay tolerance

We define the maximum delay that can be tolerated by MTC users as *delay tolerance* (DT). Even if the users request some common data, their delay requirements may be different. Corresponding to the pattern of transmission order $\boldsymbol{I}$, we define the information delay tolerance requirement matrix $\boldsymbol{D}$ as follows:

$$\boldsymbol{D} = \begin{pmatrix} d(B_1^1) & d(B_1^2) & \ldots & d(B_1^K) \\ d(B_2^1) & d(B_2^2) & \ldots & d(B_2^K) \\ \vdots & \vdots & \vdots & \ddots \\ d(B_M^1) & d(B_M^2) & \ldots & d(B_M^K) \end{pmatrix},$$

where $d(B_m^k) \in \{1, 2, 3, \ldots, K\}$ denotes the transmission deadline of data block B_m^k. Different types of users and information may have various delay tolerance. So $d(B_m^k)$ can be expressed as

$$d(B_m^k) = k + \delta_m^k, \tag{3}$$

where δ_m^k is the maximum tolerated delay of the kth data block of user m and they are i.i.d. variables that obey some truncated distributions. Here we adopt the truncated normal distribution [19]

$$f_N(\delta) = \left(\frac{\mathrm{e}^{-\frac{(\delta-\mu)^2}{2\sigma^2}}}{\int_0^{K-k} f_N^*(y)\mathrm{d}y\sqrt{2\pi}\sigma} \right), \qquad \text{for} \quad 0 \leqslant \delta \leqslant K-k, \tag{4}$$

where μ and σ are the mean and variance of delay requirement and $f_N^*(y)$ is the density function of normal distribution

$$f_N^*(y) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}}, \qquad \text{for} \quad -\infty \leqslant \delta \leqslant +\infty. \tag{5}$$

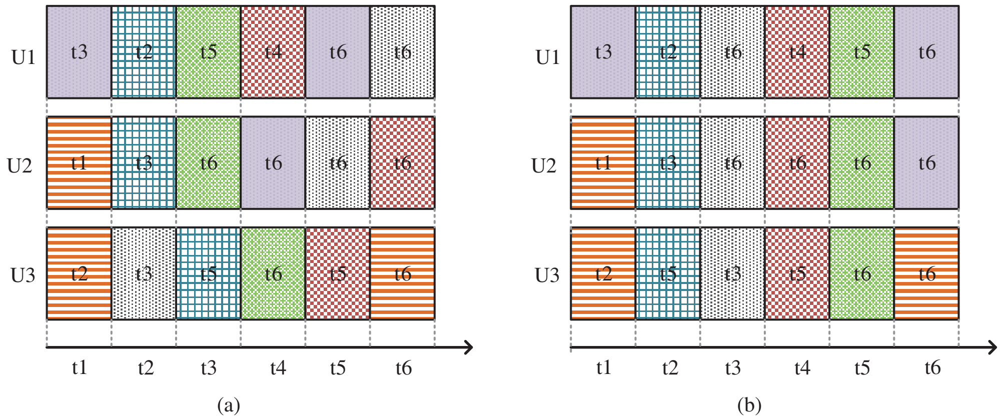

Figure 1 Demonstration of proposed scheme. (a) Before adjusting; (b) after adjusting.

3 Aggregation transmission scheme

The transmission order and delay requirements of all user data blocks in a specific time window can be known through some machine learning, statistical inference and prediction techniques based on massive wireless big data analysis. As clustering MTC users are likely to request similar or correlated data content in almost similar period of time and most of the information data transmissions may tolerate some delay, we can adjust the transmission order to make the users share the same copy of information under the delay constraints. In this way, we can save much more timeslots or some other type of radio resources.

After adjusting we can get new transmission order as

$$\widetilde{\boldsymbol{I}} = \begin{pmatrix} \widetilde{B}_1^1 & \widetilde{B}_1^2 & \ldots & \widetilde{B}_1^K \\ \widetilde{B}_2^1 & \widetilde{B}_2^2 & \ldots & \widetilde{B}_2^K \\ \vdots & \vdots & \vdots & \ddots \\ \widetilde{B}_M^1 & \widetilde{B}_M^2 & \ldots & \widetilde{B}_M^K \end{pmatrix},$$

which satisfies

$$\widetilde{\boldsymbol{D}} = \begin{pmatrix} d(\widetilde{B}_1^1) & d(\widetilde{B}_1^2) & \ldots & d(\widetilde{B}_1^K) \\ d(\widetilde{B}_2^1) & d(\widetilde{B}_2^2) & \ldots & d(\widetilde{B}_2^K) \\ \vdots & \vdots & \vdots & \ddots \\ d(\widetilde{B}_M^1) & d(\widetilde{B}_M^2) & \ldots & d(\widetilde{B}_M^K) \end{pmatrix}.$$

Figure 1 demonstrates an adjusting process of transmission order in a time division multiple access (TDMA) system. In this process, each user out of the three is requesting six data blocks each with a required transmission deadline denoted as t_i in the grids. Figure 1(a) shows the original $\boldsymbol{I}$ and $\boldsymbol{D}$ and Figure 1(b) shows the ones after adjustment. So in this case the original TDMA transmission mode needs $3 \times 6 = 18$ timeslots while our proposed ATS only occupied 8 timeslots after adjusting the transmission order.

In a TDMA system, each user will be assigned one timeslot in a frame, so we define a metric of gain ratio to evaluate the ratio of saved slot resources, namely the spectrum efficiency of ATS, as

$$R = 1 - \frac{\sum_{k=1}^{K} S_{Mk}}{MK}. \tag{6}$$

Here, M denotes the number of devices and K denotes the number of frames consisting of several timeslots, which are in accordance with the system model. S_{Mk} represents the total number of needed timeslots in kth frame for all M users. Thus, $\sum_{k=1}^{K} S_{Mk}$ is the number of actual occupied timeslots. Thereafter, the

Table 1 Permutations of transmission order and delay constraints

Transmission order	(1 3 5)	(1 5 3)	(3 1 5)	(3 5 1)	(5 1 3)	(5 3 1)
Delay constraints	(2 2 3)	(2 3 2)	(2 2 3)	(2 3 2)	(3 2 2)	(3 2 2)

optimization problem can be expressed as below

$$\begin{aligned} &\max\,(R) \\ \text{s.t.}\quad & d(\widetilde{B}_m^k) \geqslant k, \quad m \in (1, M), \quad k \in (1, K). \end{aligned} \tag{7}$$

Suppose each of M users requests K data blocks, the total number of possible permutations is $(K!)^M$. However, considering the computational load, it is infeasible to solve the problem by ergodic enumeration, especially for massive users. So some improved search algorithms for adjusting the transmission order are presented in Section 4.

4 Order optimization algorithms

Optimal transmission order can be obtained by ergodic searching algorithm, which however costs huge amount of time so that it does not satisfy the transmission requirements. Thus we come up with several order adjustment algorithms with relative low complexity. In the following parts we will expound on these algorithms.

4.1 Conditional random search

In order to reduce computation time of ergodic searching, we propose a conditional random search (CRS) scheme which is based on legal collection. We need to do some preprocessing to exclude all the solutions that do not satisfy the delay requirements. So the preprocessing of CRS has two steps: permutation generation and solution sifting. Take a simple example where one user requires 3 data blocks (1,3,5) with delay requirement (2,2,3). First we get six permutation (1 3 5), (1 5 3), (3 1 5), (3 5 1), (5 1 3) and (5 3 1), as shown in Table 1. Then taking the delay requirements into consideration, we find that only (1 3 5) and (3 1 5) satisfy the delay constraints, thus other four permutations are excluded.

After this step each user will get its own legal permutation collection and the size of each user's collection might be different because of their different delay requirements. CRS will randomly select one element from the legal collection of each user and combine all M elements to form the new solution for testing the performance. In this work we specifically define a 10000000-time searching. During all the searching rounds we always keep the best searching result. However, this algorithm still costs large amount of computing time. We will use it as a baseline for other algorithms mentioned below.

4.2 Standard-row algorithm

Instead of maximizing the IR of all users we come up with a simple algorithm in which we select a user as standard one and try to maximize IR between all other users and the standard one, as shown in Algorithm 1. The flow chart of SRA is also given as Figure 2.

First recall that B_m^k denotes the data block required by user m at timeslot k and $d(B_m^k)$ indicates the latest transmission time of data block B_m^k. Based on $\boldsymbol{D}$, we can obtain $\boldsymbol{D'}$ which indicates the acceptable adjusting range, namely $d'(B_m^k) = d(B_m^k) - k$. Then we calculate the sum of each row of $\boldsymbol{D'}$ to indicate the overall degree of adjusting range so that we can choose a user which has the most strict delay requirement as the standard user a. The information sequences of all other users are compared with user a and are adjusted to make the data blocks common with a are aligned with a as much as possible under the constraint of delay requirements. This means maximizing $Q_{u,a}$. After reordering process we can get new $\widetilde{\boldsymbol{I}}$ which has more identical data blocks at the same time and $\widetilde{\boldsymbol{D}}$. Although we can not obtain the optimal result, we can actually improve Q in some degree by enhancing $Q_{u,a}$.

Algorithm 1 Standard-row algorithm

Input: M, N, K
Output: $\widetilde{\boldsymbol{I}}$, $\widetilde{\boldsymbol{D}}$
for $m = 1, 2, \ldots, M$ **do**
  for $k = 1, 2, \ldots, K$ **do**
    Sample B_m^k
    Sample $d(B_m^k)$
$\widetilde{\boldsymbol{I}} = \boldsymbol{I}$, $\widetilde{\boldsymbol{D}} = \boldsymbol{D}$
for $m = 1, 2, \ldots, M$ **do**
  for $k = 1, 2, \ldots, K$ **do**
    $\boldsymbol{D}'[m][k] = \boldsymbol{D}[m][k] - k$
    let $\sum_{k=1}^{K} \boldsymbol{D}'[a][k]$ be the minimum
while $u < M$ **do**
    if $u \neq a$ **then**
    for $k = 1, 2, \ldots, K$ **do**
      for $kk = 1, 2, \ldots, K$ **do**
        if $\boldsymbol{I}[u][k] == \boldsymbol{I}[a][kk]$ **then**
          $interval = kk - k$
          if $0 < interval \leqslant \boldsymbol{D}'[u][k]$ **then**
            $\widetilde{\boldsymbol{I}}[u][k] = \boldsymbol{I}[u][kk]$
            $\widetilde{\boldsymbol{I}}[u][kk] = \boldsymbol{I}[u][k]$
            $\widetilde{\boldsymbol{D}}[u][k] = \boldsymbol{D}'[u][kk] + interval + k$
            $\widetilde{\boldsymbol{D}}[u][kk] = \boldsymbol{D}'[u][k] - interval + kk$
          if $interval \leqslant 0$, $|interval| \leqslant \boldsymbol{D}'[u][k]$ **then**
            $\widetilde{\boldsymbol{I}}[u][k] = \boldsymbol{I}[u][kk]$
            $\widetilde{\boldsymbol{I}}[u][kk] = \boldsymbol{I}[u][k]$
            $\widetilde{\boldsymbol{D}}[u][k] = \boldsymbol{D}'[u][kk] + interval + k$
            $\widetilde{\boldsymbol{D}}[u][kk] = \boldsymbol{D}'[u][k] - interval + kk$
      $kk \leftarrow kk + 1$
    $k \leftarrow k + 1$
  $u \leftarrow u + 1$
end while
$\widetilde{\boldsymbol{I}}$, $\widetilde{\boldsymbol{D}}$

This algorithm only focuses on maximizing the information relevance between two users instead of the whole one. So it has low complexity though the performance is not optimal.

4.3 Genetic algorithm

Genetic algorithm (GA) is an adaptive group searching method to obtain sub-optimal solution when the optimization problem is hard to solve [20]. It is used in broad fields for its implementation simplicity. So we apply GA to our optimization problem and obtain a near-optimal result.

The individual generation issue is the same as it in Subsection 4.1. Each device randomly selects one legal permutation from its collections respectively and all selected permutations will form one new individual. Eq. (6) is used to evaluate the individuals. A single numerical fitness, which is supposed to be proportional to the "utility" or "ability" of the individual will be given by the evaluation function [20]. The individuals with higher fitness are of good utility, because much more timeslot resources can be saved.

Here we adopt the integer encoding strategy for its simplicity, although the binary encoding strategy can work too. Such strategy is easy for the implementation of crossover and mutation operations, due to our problem background. The sequence identifier (ID) of each user is encoded as one gene, so that the sequence IDs of all users are lined to represent the chromosome, which stands for one possible solution. For example, if transmission order is as follows:

$$\boldsymbol{I} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

the chromosome will be formed like Figure 3.

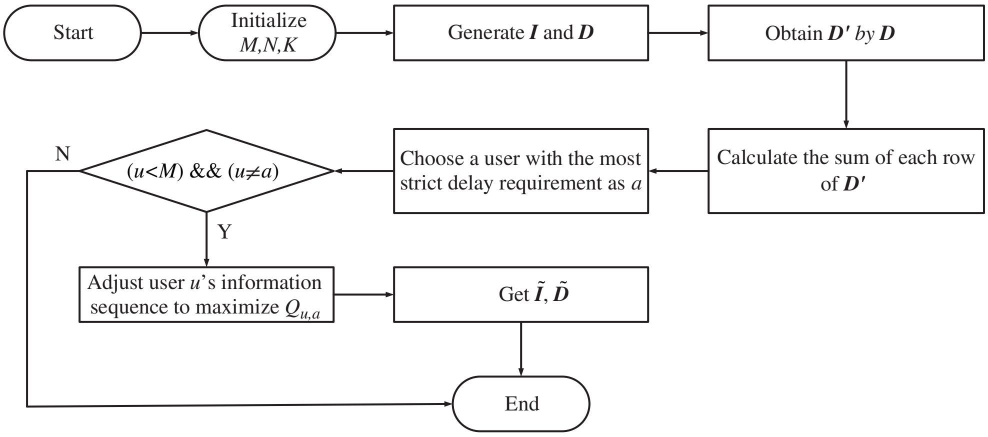

Figure 2 Flow chart for SRA.

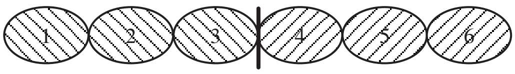

Figure 3 Example of a chromosome.

All offsprings are produced by mating two selected individuals from current population. We adopt roulette wheel selection in this work which favors the more adaptive individuals, shown as below:

$$p_c = \frac{R_c}{\sum_{c=1}^{C} R_c}, \tag{8}$$

where p_c is the probability of cth individual to be selected, R_c is the fitness of it and C is the size of population. Good individuals may be chosen several times in a generation and poor ones may not be picked up at all by this selection strategy.

After selecting two parents, we combine their chromosomes by single point crossover as shown in Figure 4. Crossover takes two individuals, and separates their chromosome strings at some randomly chosen position then exchanges the two segments of each individual. Note that we do not utilize the mutation operation in our algorithm. It is easy to know that GA costs much less computation time than the normal searching algorithm. We will see that it also has a better performance than the simple SRA.

5 Performance analysis

We will compare the performance of ATS with traditional MBMS using SRA, CRS and GA methods. We consider that the M users request part of the data from a data pool with N data blocks, so that the data of the users are correlated. We use metric of gain ratio as defined in (6) to discuss the impact of M, N and K. In addition, we also consider the influence of different delay requirements. We set in the GA algorithm that searching process stops if the optimal solution keeps unchanged for continuous 50-generations. The results in this paper are the mean value of 50-times independent searching.

5.1 Gain ratio performance

The results of comparison between ATS and traditional MBMS with different parameters are shown in Figures 5 and 6. Here, we set $K = 8$, $\mu = 2$ and $\sigma = 0.6$.

In Figure 5 for all algorithms, when N is fixed, the gain ratio is getting higher with the increasing M. We also notice that for the same M, smaller N means better performance. Actually the size of M and N reflect the degree of information relevance. Higher Q means one data block may be needed by more users, thus one timeslot resource can be shared for more users which will definitely improve the efficiency. Moreover, when the relevance is too low for $M = 32,\ N = 1024$ as shown in Figure 5, ATS is of no much help .

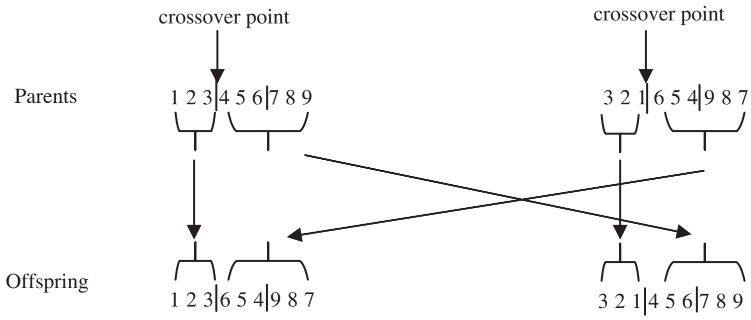

Figure 4 Single point crossover.

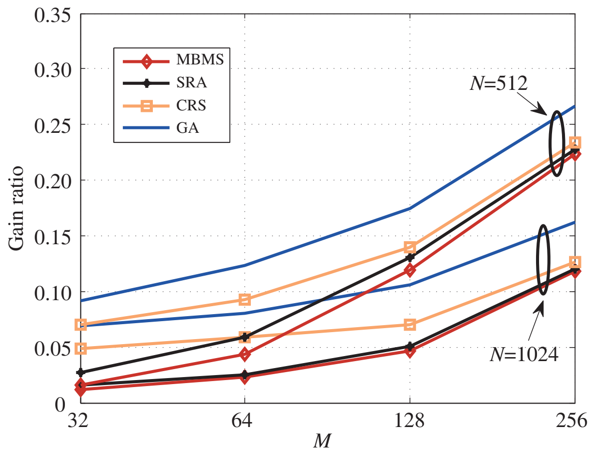

Figure 5 Performance of different algorithms with $K = 8$.

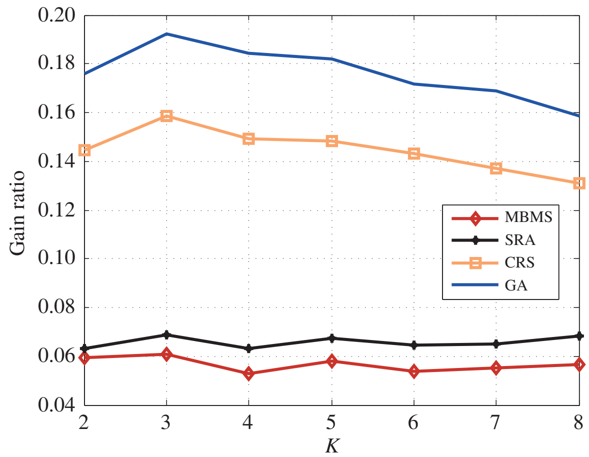

Figure 6 Performance of different algorithms with $M = 32$, $N = 256$.

In Figure 6, with the increasing number of prediction frames K, the performance of traditional MBMS keeps stable. This is because we assume the users request information in identical distribution, so the performance of traditional MBMS has no relationship with K as proved in [21]. We notice that at the beginning the performance of both CRS and GA are getting better and better with K increasing. While larger K means more calculations and higher complexity of adjustment, the performance of GA and CRS decline when K increases.

For these algorithms, computation complexity is an essential parameter. Here we use computation time in Matlab simulation systme to present the complexity as shown in Tables 2 and 3. As we can see that CRS always costs too much time although it has a better performance than SRA. With the increasing of M, both SRA and GA will cost more computation time as shown in Table 2. While GA increases more quickly than SRA which means GA is not suitable for massive users scenario. In addition, larger K always takes more calculation time, especially for GA as shown in Table 3. So on the one hand, we had better strike a balance between the energy efficiency and spectrum efficiency. On the other hand, with the increasing K, although the performance of GA is better, it has too high complexity so it can not satisfy the transmission requirements. Therefore we had better adopt SRA when K is large.

As we can see from the figures that the performance of ATS is always better than traditional MBMS which proves the validity of our proposed scheme. GA outperforms other algorithms and it is even better than the 10000000-time CRS. SRA has lowest complexity which is more suitable for larger scale optimization. CRS costs too much time in searching, but its performance can be a good reference standard of other algorithms.

Table 2 Computation time of different M, $K = 8$, $N = 512$ (unit: s)

	$M = 32$	$M = 64$	$M = 128$	$M = 256$
SRA	0.00066	0.0007	0.00084	0.0013
CRS	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$
GA	3.7660	4.5600	4.9767	6.6651

Table 3 Computation time of different K, $M = 32$, $N = 256$ (unit: s)

	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$	$K = 8$
SRA	0.0003	0.00031	0.0003	0.0003	0.00032	0.00031	0.0003	0.00033
CRS	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$	$\gg 1000$
GA	0.6249	0.9335	1.4886	2.1250	2.6140	2.8576	3.4698	3.8372

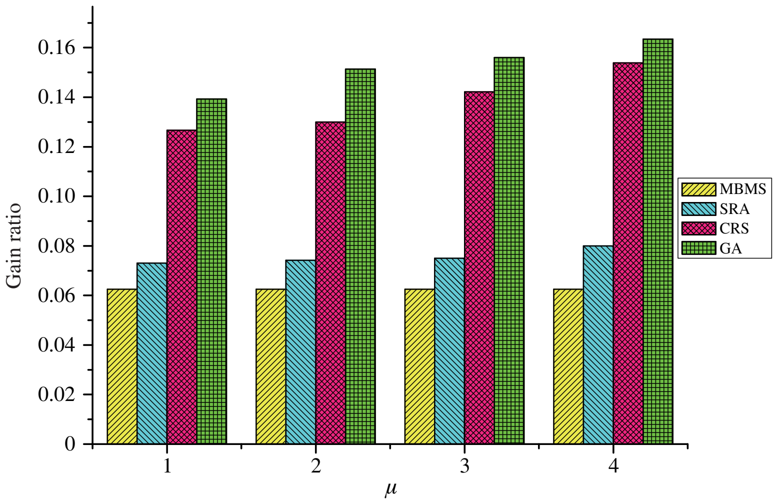

Figure 7 Gain ratio with different degree of delay requirements.

5.2 Impact of delay requirements

Different degree of delay requirements determines different adjusting space which will definitely influence the performance. By changing μ we can simulate different degree of delay requirements. μ can decide the average level of delay tolerance of all users and the higher one means looser requirements. Here, we set $M = 32$, $N = 256$ and $K = 8$, together with $\sigma = 0.6$.

The influences on the performance are shown in Figure 7. For the same original transmission order, as mentioned above larger μ means looser delay requirements which will lead to a larger adjusting space. So the performance of all algorithms are all improved with the increasing μ. Because of its limited ability, SRA gets better slightly. The performances of CRS and GA are improved more obviously. While we also notice that larger μ may consume more computing time, especially in the process of generating legal collection.

6 Conclusion

Aiming at aggregating MTC users which are likely to request correlated contents almost at the same time in one cell, we proposed ATS to improve the transmission efficiency. By fully using prediction techniques and users' delay tolerance, ATS allows the adjustment of transmission order. We compared the performance of ATS with traditional MBMS using three different algorithms, namely CRS, SRA and GA. The results validated the effectiveness of our proposed ATS. Moreover, the performance of GA outperforms other algorithms and SRA has the lowest complexity which is more suitable for large scale adjustment. The impact of different degree of delay requirements was also evaluated through simulation.

Acknowledgements This work was partially supported by Natural Science Foundation of China (Grant No. 61461136002), Key Program of National Natural Science Foundation of China (Grant No. 61631018), Fundamental Research Funds for the Central Universities, and Huawei Innovation Research Program.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Chen K C. Machine-to-machine communications for healthcare. J Comput Inf Sci Eng, 2012, 6: 119–126
2 Fadlullah Z M, Fouda M M, Kato N, et al. Toward intelligent machine-to-machine communications in smart grid. IEEE Commun Mag, 2011, 49: 60–65
3 Wu G, Talwar S, Johnsson K, et al. M2M: from mobile to embedded Internet. IEEE Commun Mag, 2011, 49: 36–43
4 Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021. White Paper. 2017
5 Ericsson L. More than 50 billion connected devices. White Paper. 2011
6 Dawy Z, Saad W, Ghosh A, et al. Toward massive machine type cellular communications. IEEE Wirel Commun, 2017, 24: 120–128
7 Shariatmadari H, Ratasuk R, Iraji S, et al. Machine-type communications: current status and future perspectives toward 5G systems. IEEE Commun Mag, 2015, 53: 10–17
8 Ma Z, Zhang Z Q, Ding Z G, et al. Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives. Sci China Inf Sci, 2015, 58: 041301
9 Andrews J G, Buzzi S, Choi W, et al. What will 5G be? IEEE J Sel Areas Commun, 2014, 32: 1065–1082
10 Chen X, Chen T Y, Guo D N. Capacity of Gaussian many-access channels. IEEE Trans Inf Theory, 2017, 63: 3516–3539
11 Xie R G, Yin H R, Chen X H, et al. Many access for small packets based on precoding and sparsity-aware recovery. IEEE Trans Commun, 2016, 64: 4680–4694
12 Zhang G P, Liu P, Yang K, et al. Orthogonal resource sharing scheme for device-to-device communication overlaying cellular networks: a cooperative relay based approach. Sci China Inf Sci, 2015, 58: 102301
13 Lien S Y, Chen K C, Lin Y. Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. IEEE Commun Mag, 2011, 49: 66–74
14 Lecompte D, Gabin F. Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements. IEEE Commun Mag, 2012, 50: 68–74
15 Valerio D, Ricciato F, Belanovic P, et al. UMTS on the road: Broadcasting intelligent road safety information via MBMS. In: Proceedings of the 2008 IEEE 67th Vehicular Technology Conference (VTC-Spring), Singapore, 2008. 3026–3030
16 Chin W H, Fan Z, Haines R. Emerging technologies and research challenges for 5G wireless networks. IEEE Wirel Commun, 2014, 21: 106–112
17 Lu X, Wetter E, Bharti N, et al. Approaching the limit of predictability in human mobility. Sci Rep, 2013, 3: 1–9
18 Song C M, Qu Z H, Blumm N, et al. Limits of predictability in human mobility. Science, 2010, 327: 1018–1021
19 Hung H N, Lin Y B, Lu M K, et al. A statistic approach for deriving the short message transmission delay distributions. IEEE Trans Wirel Commun, 2004, 3: 2345–2352
20 Beasley D, Bull D R, Martin R R. An overview of genetic algorithms: part 1, fundamentals. Univ Comput, 1993, 15: 56–69
21 Sun Y H, Zhang S H, Zhu J K, et al. Aggregation postpone transmission scheme for machine type communications. In: Proceedings of the 2016 IEEE 19th Wireless Personal Multimedia Communications (WPMC), Shenzhen, 2016. 205–210