

面向 mMTC 的 5G 网络多随机接入机制性能优化策略

詹文¹, 梁译文¹, 孙兴华¹, 谭雪松²

(1. 中山大学·深圳电子与通信工程学院, 广东 深圳 518107;

2. 电子科技大学通信抗干扰技术国家级重点实验室, 四川 成都 611731)

摘要: 为优化对具有不同流量特征的海量机器类通信 (mMTC, massive machine type communication) 业务的支持, 5G 网络将并存两种随机接入机制: 基于报文的随机接入 (PBRA, packet-based random access) 机制和基于连接的随机接入 (CBRA, connection-based random access) 机制。然而, 面向 mMTC 业务流量特征的随机接入机制选择问题, 却尚未得到解决。为此, 本文将分析 PBRA 机制和 CBRA 机制下的网络吞吐量和信令开销, 解决在信令开销受限的条件下, 吞吐量最大化的优化问题。以此为基础, 得到了最优随机接入机制选择策略和相应的网络参数配置方案, 并通过仿真进行验证。

关键词: mMTC; 基于报文的随机接入; 基于连接的随机接入; 随机接入机制选择

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2022.00267

Performance optimization for mMTC in 5G network with multiple random access schemes

ZHAN Wen¹, LIANG Yiwen¹, SUN Xinghua¹, TAN Xuesong²

1. School of Electronics and Communication Engineering, Sun Yat-sen University (Shenzhen), Shenzhen 518107, China

2. National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: To optimize the support for massive machine type communication (mMTC) services with different traffic characteristics, packet-based random access (PBRA) scheme coexists with the connection-based random access (CBRA) scheme in 5G networks. Yet, given the traffic characteristic of mMTC, which random access scheme should be chosen is still an open issue. To address this issue, the network throughput and signaling overhead analysis of mMTC with PBRA and CBRA were presented, and the problem of throughput maximization while maintaining the signaling-to-throughput ratio below a certain level was solved. Based on this, the optimal access scheme selection strategy and the corresponding optimal network parameter configuration were obtained and verified via simulations.

Key words: mMTC, packet-based random access, connection-based random access, random access scheme selection

0 引言

物联网, 作为新一代信息技术的代表, 是推动

国家基础性行业发展和消费性行业升级的主要力量。作为物联网的重要组成部分, 海量机器类通信 (mMTC, massive machine type communication)

收稿日期: 2021-12-12; 修回日期: 2022-04-05

通信作者: 詹文, zhanw6@mail.sysu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62001524); 广东省重点领域研发计划 (No. 2020B0101120003); 广东省基础与应用基础研究基金资助项目 (No. 2019A1515011906); 深圳市高等院校稳定支持计划项目 (No. 2021A04); 深圳市科技计划资助 (No. RCBS20210706092408010)

Foundation Items: The National Natural Science Foundation of China (No.62001524), The Key-Area Research and Development Program of Guangdong Province (No.2020B0101120003), The Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515011906), The Shenzhen Sustainable Support Program for Universities (No.2021A04), The Shenzhen Science and Technology Program (No.RCBS20210706092408010)

是指在机器上通过安装传感器、控制器等方式赋予机器“智能”的属性,使得机器与机器间能在无人控制的情况下,自主地进行数据交互的技术。近年来大规模发展的物联网新兴产业,如智慧水务、智慧监控和可穿戴智能设备等,都需要 mMTC 的支持^[1]。预计到 2023 年,机器类设备(MTD, machine type device)的数量将超过 140 亿^[2]。由于其巨大的市场潜力, mMTC 已被视为 5G 系统的三大通用服务场景之一。

蜂窝系统长久以来都采用基于连接的随机接入(CBRA, connection-based random access)方案,即每个设备在传输数据之前,必须执行随机接入过程以与基站(在 5G 系统中称为 gNB)建立连接,然后再与基站传输数据^[3]。CBRA 方案旨在支持传统的以人为中心的通信(HTC, human-type communication),如视频流等。在 HTC 的应用场景下, HTC 设备数量相对较少,但每个设备需传输大量数据。因此,建立连接过程所产生的信令开销可以忽略不计。

然而, mMTC 的流量特征却完全不同于 HTC。在 mMTC 的应用场景下, MTD 数量庞大,并且以上行的小数据报文传输为主,例如,传感器发送的数据报文可能仅十几比特大小^[4]。因此,在 CBRA 机制下,建立连接过程所产生的信令开销无法忽视,对于 MTD 来说,相对较多的信令开销会导致额外的时延,加快 MTD 电池电量消耗速度^[5]。为了解决信令问题,5G 标准引入了基于报文的随机接入(PBRA, packet-based random access)方案,即 MTD 可以在随机接入过程中传输一个小数据包,而无须与基站建立连接^[6]。文献[7]和文献[8]中的仿真结果表明,与 CBRA 方案相比,PBRA 方案可以显著降低能耗、时延和信令开销。

尽管 PBRA 方案是在 5G 网络中支持 mMTC 的一种有效机制,但其性能依然取决于网络参数配置^[9]。主要的原因在于 PBRA 是基于 Aloha 协议设计的,即每个 MTD 独立地确定何时发送请求。这种分布式行为意味着在大规模接入场景下,可能会出现大量 MTD 同时发起接入请求的情况,导致严重的冲突问题和极低的成功接入机率,使得 MTD 频繁进行数据包重传,引发巨大的信令开销。因此,研究如何调整 5G 网络 PBRA 方案中的系统参数以优化接入效率,并进一步降低信令开销具有重要现实意义。

在已有的工作中,对 Aloha 型接入网络的研究

集中在吞吐量性能上,即每个时隙成功传输的数据包或比特数的长期平均数。例如,基于对信道负载的周期性估计,文献[10-16]提出了多种策略动态地配置 MTD 的传输概率,以降低冲突概率并最大化吞吐量。在给定随机接入信道上的接入请求的流量服从泊松分布或者单个节点内数据报文到达过程服从伯努利分布的情况下,文献[17-20]提出了多类分析框架以刻画每个设备或设备内队首(HOL, head of line)数据包的行为,获得稳态下成功传输概率、吞吐量的显式表达式,进而求解得到吞吐量最大化时的最佳退避参数设置方案。然而以上的工作均未考虑信令开销。针对 PBRA 和 CBRA 性能优化的一些关键问题仍未得到解决,例如,吞吐量最大化是否等同于信令开销最小化?如果为信令开销设置上限要求,那么又该如何调整系统参数来优化吞吐量性能?

另一方面,随着物联网应用的不断深化, mMTC 的流量特征(如节点数目、报文到达速率和报文大小等)随业务场景的变化而动态变化。不同的随机接入机制适用的 mMTC 业务场景不同。CBRA 机制通常适用于 mMTC 节点频繁传输数据报文或者传输相对较大的数据报文的场景,而 PBRA 机制适用于 mMTC 节点非频繁地传输小数据报文的场景,当随机接入机制选择不恰当时,例如,原本适合于 CBRA 机制的 mMTC 业务使用了 PBRA 机制,那么 MTD 频繁发送接入请求和传输数据报文,会造成网络拥塞和过量信令开销的问题。因此,有必要为不同的 mMTC 业务选择合适的随机接入机制。然而,现在鲜有研究关注随机接入机制的选择问题。文献[21]虽然考虑了该问题,但是其忽视了 CBRA 机制和 PBRA 机制在接入过程中信令开销的差异性,也未考虑有服务质量要求下的吞吐量优化策略。文献[22]用固定的数据包大小作为随机接入机制选择的判决门限,缺乏动态性。因此,在 CBRA 方案和 PBRA 方案共存的 5G 网络中,如何根据 mMTC 的流量特征动态地选择随机接入机制是尚待解决的问题。

本文旨在为上述开放性问题的提出解决方案,通过信令与吞吐量的比率(称其为信吞比)评估信令开销,它表示每成功传输 1 bit 所需要的平均信令开销。本文贡献如下。

在 PBRA 机制下,本文刻画了网络在饱和/不饱和情况下的吞吐量和信吞比性能,解决了在信吞比

低于上限 β 的限制条件下的吞吐量最优化问题，得到了最大吞吐量和最优接入概率（也就是接入类禁止（ACB, access class barring）因子^[23]）的显式表达式。本文还进一步将研究拓展至 CBRA 场景，刻画了网络在不饱和情况下的最优吞吐量和信吞比性能。

基于上述分析结果，本文以吞吐量最大化为目标，以信吞比低于上限 β 为限制条件，提出了给定 MTD 数目 n 、数据到达率 λ 、数据传输速率 D 和数据报文长度 L 情况下的动态随机接入机制选择与参数配置方案。本文的研究显示，随 mMTC 业务数据量的增加，网络应该先选择 PBRA 机制，随后切换到 CBRA 机制。在过高的数据量下，网络反而需要切换回 PBRA 机制。本文还阐释在 5G 网络中应用本文所提出的随机接入机制选择方案的方法，并通过仿真验证了在该方案下，5G 网络可以更好地支撑 mMTC 业务。

1 系统模型

考虑单一小区的 5G 系统，基站（gNB）为 n 个 MTD 提供服务。每个 MTD 的数据包到达率服从参数为 λ 的伯努利过程，并假定 MTD 的数据缓存队列空间无限大。当有数据在缓存队列并且 MTD 与基站没有数据连接时，MTD 就会发起 4 步随机接入过程。

在当前的 5G 系统中，PBRA 方案和 CBRA 方案共存，5G 网络中共存的两种随机接入方案如图 1 所示。本文将同时考虑两个方案。

如图 1(a)所示，在 PBRA 方案中，MTD 发起 4 步随机接入过程。在第 1 步中，MTD 将从 $M(M \geq 1)$

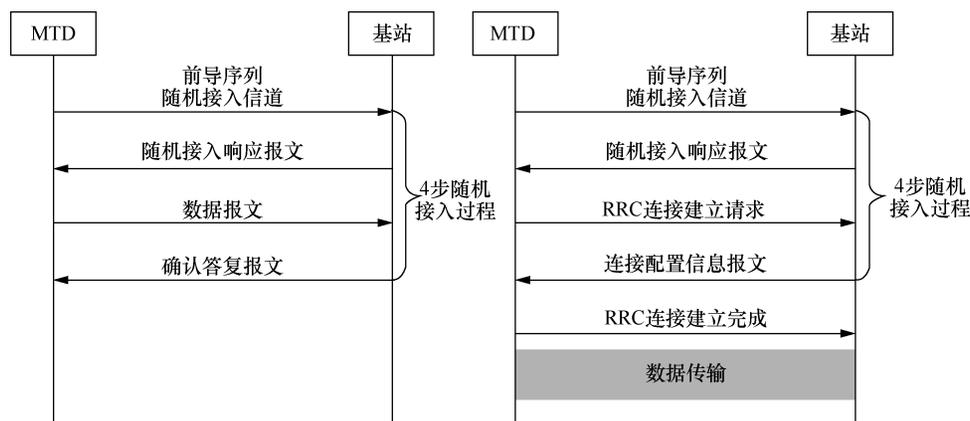
个正交前导码中随机选择一个前导码，并通过物理随机接入信道发送给基站；基站在第 2 步根据其收到的每个前导码，回复随机接入响应报文；MTD 在第 3 步传输一个数据报文；基站在第 4 步回复确认答复报文。我们假设当且仅当没有其他 MTD 在同一时隙向基站发送相同的前导码时，MTD 才能成功传输数据报文，否则，碰撞发生，传输失败。由于前导码相互正交，不同 MTD 选择不同前导码将不会影响彼此成功接入网络的机会。为了简单起见，本文仅考虑单一前导码的情况，即 $M=1$ ，而多前导码的情况可以根据文献[24]中的多组模型进行扩展。

如图 1(b)所示，在 CBRA 方案中，MTD 同样发起 4 步随机接入过程。不同的是，设备在第 3 步发送连接建立请求，并在随后的数据连接中，通过基站集中分配的独占的上行时频资源，无冲突地传输数据。假设 MTD 在每个时隙中能够传输数据包的数量为 $D(D \geq 1)$ ，即 D 为数据传输率。

假设一个时隙的长度对应 MTD 与基站间完成一次 4 步随机接入过程的时间长度。在本文中，我们还假设 5G 系统使用 ACB 方案用于接入控制，即缓存队列非空的 MTD 在每个时隙以概率 $q \in (0, 1]$ 向基站发送接入请求，也就是前导码。在 5G 标准里面，传输概率 q 也被称为 ACB 因子^[23]。

1.1 性能指标

由于在 PBRA 方案中每个数据包都要竞争信道，而在 CBRA 方案中，只有缓存队列中的 HOL 数据包需要竞争信道，因此，不同随机接入方案中的平均信令开销会存在差异。本文将通过吞吐量和信令开销这两个性能指标来研究并解决随机接入



(a) 基于报文的随机接入方案

(b) 基于连接的随机接入方案

图 1 5G 网络中共存的两种随机接入方案

方案的选择问题。

用 $\hat{\gamma}_{\text{out}}$ 表示吞吐量，定义其为每个时隙成功传输的比特数的长期平均值。 $\hat{\gamma}_{\text{out}} = \hat{\lambda}_{\text{out}} L$ ，其中， $\hat{\lambda}_{\text{out}}$ 为数据吞吐量，定义为每个时隙成功传输的数据包数量的长期平均数， L 是每个数据包的比特数。在 PBRA 方案中，由于设备每次随机接入传输一个数据报文，所以接入吞吐量即数据吞吐量，即 $\hat{\lambda}_{\text{out}} = \hat{\lambda}_{\text{out}}^a$ ，其中接入吞吐量 $\hat{\lambda}_{\text{out}}^a$ 表示平均每时隙成功传输的接入请求数量；在 CBRA 方案中，数据吞吐量即 $\hat{\lambda}_{\text{out}}$ 。

为了表征在随机接入过程中的信令开销 S ，另外定义 s （单位为比特）表示 MTD 和 gNB 之间交换的每个控制信令的平均大小。根据 4 步随机接入是否成功，可以分为以下两种情况。

1) 如果 MTD 的 4 步随机接入失败，则信令开销为 $4s$ 。传输失败导致的长期信令开销可以写为 $4sF$ ，其中 F 表示平均每个时隙中失败的接入请求数量。

2) 如果 MTD 成功完成随机接入过程，则在 PBRA 方案中，如图 1 (a) 所示，由于在随机接入第 3 步传输的是数据包，其不计入信令开销，所以成功传输的长期信令开销为 $3s\hat{\lambda}_{\text{out}}^a$ ；在 CBRA 方案中，由于需要与基站建立连接之后才能传输数据，如图 1(b) 所示，其信令开销为 $5s$ ，所以成功传输的长期信令开销为 $5s\hat{\lambda}_{\text{out}}^a$ 。

在此，我们可以总结在 PBRA 方案中的平均信令开销为

$$S_{\text{PBRA}} = 4sF + 3s\hat{\lambda}_{\text{out}}^a \quad (1)$$

在 CBRA 方案中的平均信令开销为

$$S_{\text{CBRA}} = 4sF + 5s\hat{\lambda}_{\text{out}}^a \quad (2)$$

1.2 优化问题

mMTC 服务期望网络吞吐量能够尽可能大，同时保证较低的信吞比 $\frac{S}{L\hat{\lambda}_{\text{out}}}$ ，即每成功传输 1 bit 所需要的平均信令开销。令 β 表示信吞比的上限，令 $\hat{\gamma}_{\text{max}}$ 表示最大网络吞吐量。本文将首先解决在给定 PBRA 或 CBRA 机制下，如何通过调整 ACB 因子 q 来最大化网络吞吐量，并保持信吞比在门限 β 之下的优化问题，即

$$\begin{aligned} \hat{\gamma}_{\text{max}} &= \max_{0 < q \leq 1} \hat{\gamma}_{\text{out}} \\ \text{s.t. } \frac{S}{L\hat{\lambda}_{\text{out}}} &\leq \beta \end{aligned} \quad (3)$$

在 PBRA 方案中 $S = S_{\text{PBRA}}$ ，在 CBRA 方案中 $S = S_{\text{CBRA}}$ 。然后再基于优化问题的解，进一步解决随机接入机制的选择问题。

2 基于报文的随机接入机制的性能优化

2.1 面向 PBRA 的马尔可夫链

为了刻画每个数据包在随机接入信道中的行为，本文使用文献[20]中的分析模型，PBRA 方案中访问请求的状态转移图如图 2 所示，通过马尔可夫链 $\mathbf{X} = \{X_j\}$ 描述 HOL 报文的状态：成功传输状态（状态 T）和等待传输状态（状态 0）。

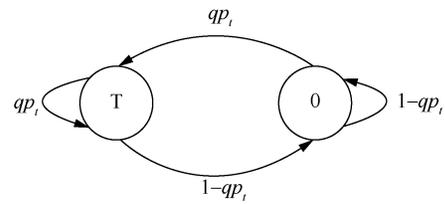


图 2 PBRA 方案中访问请求的状态转移图

可以根据文献[25]推导出图 2 中马尔可夫的稳态概率分布为 $\pi_T = pq$ 和 $\pi_0 = 1 - pq$ ，其中 $p = \lim_{i \rightarrow \infty} p_i$ 为成功传输数据包的稳态概率。 π_T 为缓存队列的服务速率。每个缓存队列非空概率可以表示为

$$\rho = \frac{\lambda}{\pi_T} = \frac{\lambda}{pq} \quad (4)$$

对于每个 HOL 数据包，当且仅当其余 $n-1$ 个节点以概率 $1-\rho$ 处于空闲状态，或以概率 $\rho(\pi_0 + \pi_T)(1-q)$ 处于队列非空但不请求接入的状态时，该数据包才能成功传输。根据非空概率 ρ 是否等于 1，可以将网络分为网络饱和/不饱和情况。当 $\rho=1$ 时，网络处于饱和情况；当 $\rho < 1$ 时，网络处于不饱和情况。

2.2 网络饱和情况下的性能优化

当 $\rho=1$ 时，网络处于饱和情况，MTD 的缓存队列一直非空，所以每个 MTD 在每个时隙都会以概率 q 发起随机接入过程。此时，HOL 数据包的成功传输概率为 $p_A = (1-q)^{n-1} \approx \exp(-nq)$ 。每个时隙中传输失败的数据包平均数量可以表示为

$$F = nq(1 - p_A) \quad (5)$$

在饱和情况下，数据吞吐量由服务速率决定，即

$$\hat{\lambda}_{\text{out}} = n\pi_T = n \exp(-nq)q \quad (6)$$

结合式(1)、式(5)、式(6), 优化问题(3)可以表示为

$$\hat{\gamma}_{\max}^{\text{sa}} = \max_{0 < q \leq 1} Ln \exp(-nq)q \quad (7)$$

$$\text{s.t. } \frac{4s \exp(nq) - s}{L} \leq \beta$$

其中, $\hat{\gamma}_{\max}^{\text{sa}}$ 为饱和情况下的最大网络吞吐量。通过求解上述优化问题, 得到在饱和情况下的最大吞吐量为

$$\hat{\gamma}_{\max}^{\text{sa}} = \begin{cases} Le^{-1}, & \frac{(4e-1)s}{L} \leq \beta \\ \frac{4Ls}{\beta L + s} \ln \frac{\beta L + s}{4s}, & \frac{3s}{L} < \beta < \frac{(4e-1)s}{L} \\ \text{无解}, & 0 < \beta \leq \frac{3s}{L} \end{cases} \quad (8)$$

并且对应的最佳 ACB 因子 q 为

$$q_{\text{sa}}^* = \begin{cases} \frac{1}{n}, & \frac{(4e-1)s}{L} \leq \beta \\ \frac{1}{n} \ln \frac{\beta L + s}{4s}, & \frac{3s}{L} < \beta < \frac{(4e-1)s}{L} \\ \text{无解}, & 0 < \beta \leq \frac{3s}{L} \end{cases} \quad (9)$$

其中, q_{sa}^* 为饱和情况下使得网络吞吐量最大化的 ACB 最优值。在理想情况下, 所有数据包都可以一次性成功传输, 信令吞吐量为 $\frac{3s}{L}$ 。因此, 如果该信吞比上限 $\beta \leq \frac{3s}{L}$, 则式(7)无解。

2.3 网络不饱和情况下的性能优化

当用户终端的数量或流量输入速率降低时, 网络将变成不饱和情况, 即每个节点缓存队列非空概率 $\rho < 1$ 。此时, 数据包传输成功的概率为 $p = (1 - \rho + \rho(1 - q))^{n-1} \approx \exp\left(-\frac{\hat{\lambda}}{p}\right)$, 根据文献[20]得到稳态运行点 $p_L = \exp(-\mathbb{W}_0(-\hat{\lambda}))$, 其中 $\hat{\lambda}$ 为总的到达率 ($\hat{\lambda} = n\lambda$)。根据式(4)和 p_L , 每个时隙的数据包传输失败的平均数量可以表示为

$$F = n\rho q(1 - p_L) = \hat{\lambda} \left(\exp(-\mathbb{W}_0(-\hat{\lambda})) - 1 \right) \quad (10)$$

在不饱和的情况下, 网络是稳定的, 即数据吞吐量等于总的数据输入的速率, 也就是 $\hat{\lambda}_{\text{out}} = \hat{\lambda}$ 。联合式(1)和式(10), 优化问题式(3)可以表示为

$$\hat{\gamma}_{\max}^{\text{unsa}} = \max_{q \in S_L} L\hat{\lambda} \quad (11)$$

$$\text{s.t. } \frac{4s \exp(-\mathbb{W}_0(-\hat{\lambda})) - s}{L} \leq \beta$$

其中, $\hat{\gamma}_{\max}^{\text{unsa}}$ 为不饱和情况下的最大网络吞吐量。通过求解上述优化问题, 可以得到在饱和情况下的最大吞吐量为

$$\hat{\gamma}_{\max}^{\text{unsa}} = \begin{cases} L\hat{\lambda}, & \frac{4s \exp(-\mathbb{W}_0(-\hat{\lambda})) - s}{L} \leq \beta \\ \text{无解}, & \frac{4s \exp(-\mathbb{W}_0(-\hat{\lambda})) - s}{L} > \beta \end{cases} \quad (12)$$

并且对应的最佳 ACB 因子 q 为

$$q_{\text{unsa}}^* \in S_L = [q_l, q_u] = \begin{cases} \left[\frac{\hat{\lambda}}{np_L}, -\frac{1}{n} \ln p_s \right], & n\lambda < e^{-1} \\ \emptyset, & n\lambda \geq e^{-1} \end{cases} \quad (13)$$

其中, q_{unsa}^* 为不饱和情况下使得网络吞吐量最大化的 ACB 最优值。式(13)中的 S_L 为绝对稳定区域^[20], 在此区域内, 网络是不饱和的并且运行在稳态点 p_L 。从式(11)可以看出, 网络吞吐量和信吞比都与 ACB 因子 q 无关, 这意味着, q 可以从绝对稳定区域 S_L 内任意取值。

2.4 性能优化总结

通过前面的分析, 得到了在网络饱和以及不饱和的情况下, PBRA 机制下最优的 ACB 参数配置方案, 在最大化吞吐量的同时满足信吞比的要求, 文献[26]已通过仿真验证以上分析。为了综合上述分析, 同时也为了方便 5G 系统在给定网络场景下, 迅速判断出 PBRA 机制下最优 ACB 参数的取值, 本文进一步以 $\{n, \lambda, \beta, L\}$ 为参数空间, 定义 PBRA 中的最大吞吐量以及最优的 ACB 因子, 见表 1, 其中区域 S_1 和 S_2 属于不饱和的情况, 区域 S_3 、 S_4 和 S_5 属于饱和的情况。

1) 当 $\{n, \lambda, \beta, L\} \in S_1$ 或者 $\{n, \lambda, \beta, L\} \in S_3$ 时, 网络可以在信吞比约束条件下, 实现最大吞吐量 (在不饱和情况下为 $L\hat{\lambda}$, 在饱和情况下为 Le^{-1})。因此, 区域 S_1 和 S_3 被称为期望区域。

2) 当 $\{n, \lambda, \beta, L\} \in S_4$ 时, 有 $\frac{4Ls}{\beta L + s} \ln \frac{\beta L + s}{4s} < Le^{-1}$, 也就是说网络需要通过牺牲最大吞吐量来满足对信吞比的约束。因此, 区域 S_4 被称为非期望区域。

表 1 PBRA 中的最大吞吐量以及最优的 ACB 因子

区域	$\{n, \lambda\}$	$\{\beta, L\}$	q^*	$\hat{\gamma}_{\max}$
S_1	$n\lambda < e^{-1}$	$\frac{4s \exp(-W_0(-\hat{\lambda})) - s}{L} \leq \beta$	$\left[\frac{\hat{\lambda}}{np_L}, -\frac{1}{n} \ln p_s \right]$	$L\hat{\lambda}$
S_2		$\frac{4s \exp(-W_0(-\hat{\lambda})) - s}{L} > \beta$	无解	无解
S_3	$n\lambda \geq e^{-1}$	$\frac{(4e-1)s}{L} \leq \beta$	$\frac{1}{n}$	Le^{-1}
S_4		$\frac{3s}{L} < \beta < \frac{(4e-1)s}{L}$	$\frac{1}{n} \ln \frac{\beta L + s}{4s}$	$\frac{4Ls}{\beta L + s} \ln \frac{\beta L + s}{4s}$
S_5		$0 < \beta \leq \frac{3s}{L}$	无解	无解

3) 当 $\{n, \lambda, \beta, L\} \in S_2$ 或者 $\{n, \lambda, \beta, L\} \in S_5$ 时, 优化问题式(3)无解, 这意味着无论 q 如何取值, 都不能满足信吞比的约束条件。因此, 区域 S_2 和 S_5 被称为无解区域。

为了进一步了解区域 $S_i (i \in \{1, 2, 3, 4, 5\})$ 随系统参数的变化, 关于数据报文长度 L 的区域划分示意图如图 3 所示, 图 3 (a) 和图 3 (b) 分别给出了饱和情况下区域 S_3 、 S_4 和 S_5 以及不饱和情况下区域 S_1 和 S_2 随数据包长度 L 和信吞比的阈值 β 的变化情况。从图 3 (a) 中可以看到, 随着数据包长度 L 增大, 期望区域 S_3 迅速扩张, 而非期望区域 S_4 和无解区域 S_5 缩小。在图 3 (b) 中也可以得到类似的观察结果。这表明在使用 PBRA 方案时, 增加数据包长度可以提升网络性能。然而, 在实际的 5G 系统中, 由于上行时频资源有限, 随机接入过程的第 3 步可以携带的数据包长度也是有限的。另一方面, 图 3 还表明, 信吞比的阈值 β 的增加也可以扩大期望区域 S_1 和 S_3 。

3 基于连接的随机接入机制的性能优化

3.1 面向 CBRA 的马尔可夫更新过程

本节将把分析拓展到 CBRA 方案, 并采用文献[24]中的分析模型去表征每个 MTD 的接入请求的行为, 建立的 CBRA 方案中访问请求的状态转移图如图 4 所示, 图 4 中 3 个状态间的转移即为离散时间马尔可夫更新过程 $(X, V) = \{(X_j, V_j), j = 0, 1, \dots\}$ 。 X_j 表示在第 j 次转移时访问请求的状态, V_j 表示第 j 次转移发生的时间点。与图 2 不同的是, 图 4 中的 HOL 数据包有一个新状态, 即状态 H, 代表成功接入后的数据传输状态。令 D 表示数据传输速率, 即每个时隙中可以传输的数据包的数量。

文献[24]的研究表明如果总输入速率小于数据传输速率, 即 $\hat{\lambda} < D$, 网络在不饱和情况下运行, 并且处于稳定状态; 如果 $\hat{\lambda} \geq D$, 网络会转到饱和情况, 也就是缓存队列中的数据包的量将随时间推移趋于无穷大。在网络饱和情况下, 接入成功的 MTD

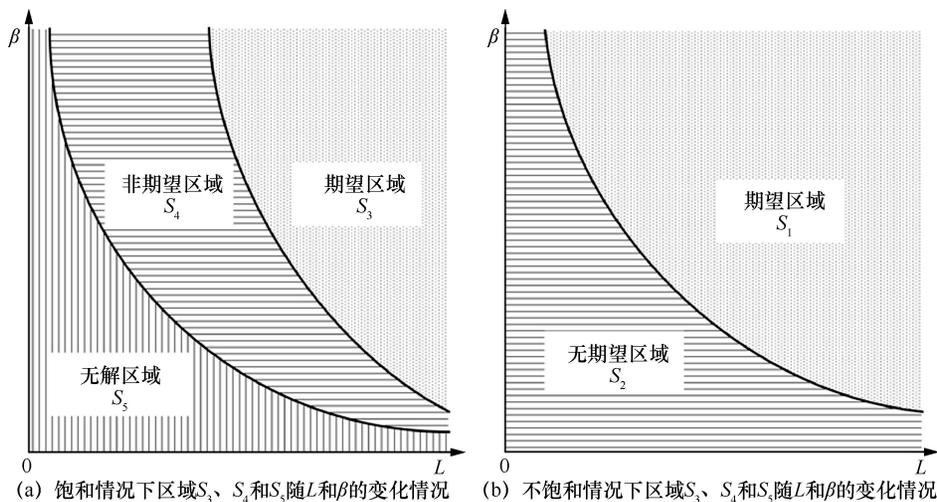


图 3 关于数据报文长度 L 的区域划分示意图

会一直占据数据信道传输数据, 而其他 MTD 无法接入, 导致严重的公平性问题^[24]。因此, 本文仅关注 $\hat{\lambda} < D$ 的情况。

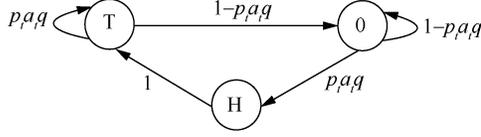


图4 CBRA 方案中访问请求的状态转移图

3.2 网络不饱和情况下的性能优化

当总输入速率 $\hat{\lambda} < D$ 时, 文献[24]推导了最大接入吞吐量的表达式

$$\hat{\lambda}_{\max}^a = \frac{nD(D - \hat{\lambda})}{D(enD - n - 1) + \frac{\hat{\lambda}D(ne - 1)}{n} + \frac{\hat{\lambda}^2}{n}} \quad (14)$$

并且对应的最优 ACB 因子设置为

$$q^* = \frac{\hat{\lambda}(D - e^{-1})}{nD(\hat{\lambda} - e^{-1})} \quad (15)$$

另一方面, 在稳定状态下, 网络的数据吞吐量等于总输入速率, 即

$$\hat{\lambda}_{\text{out}} = \hat{\lambda} \quad (16)$$

根据式(16), 网络吞吐量 $\hat{\gamma}_{\text{out}} = L\hat{\lambda}_{\text{out}} = L\hat{\lambda}$, 与 ACB 因子 q 无关, 仅仅与 MTD 数量 n 、网络输入速率 λ 和数据包的长度 L 有关。因此, 我们不妨令 $q = q^*$, 使得接入吞吐量最大化, 也就是 $\hat{\lambda}_{\text{out}}^a = \hat{\lambda}_{\max}^a$ 。于是, 优化问题式(3)可以表示为

$$\begin{aligned} \hat{\gamma}_{\max} &= \max_{0 < q \leq 1} L\hat{\lambda} \\ \text{s.t.} \quad & \frac{S}{L\hat{\lambda}} \leq \beta \end{aligned} \quad (17)$$

其中, 信令开销为

$$\begin{aligned} S &= 4sn\rho q^*(1 - p_L) + 5s\hat{\lambda}_{\max}^a = \\ & \frac{4sn\rho\hat{\lambda}(D - e^{-1})(1 - p_L)}{nD(\hat{\lambda} - e^{-1})} + \\ & \frac{5sn^2D(D - \hat{\lambda})}{nD(enD - n - 1) + \hat{\lambda}D(ne - 1) + \hat{\lambda}^2} \end{aligned} \quad (18)$$

其中, ρ 和 p_L 可以分别由文献[24]的式(6)和式(10)给出。从式(18)可以看出, 信令开销是由系统参数 n 、

λ 和 D 决定的。此时, 只要约束条件 $\frac{S}{L\hat{\lambda}} \leq \beta$ 满足, q^* 即最佳的 ACB 因子, 对应的最大网络吞吐量为 $\hat{\gamma}_{\max} = L\hat{\lambda}$ 。

4 面向 mMTC 的 5G 网络多随机接入机制选择方案

上述分析解决了在 PBRA 机制与 CBRA 机制下, 如何通过调整 ACB 因子使得系统在满足信吞比限制的情况下, 最大化吞吐量的问题。这一节, 将进一步解决随机接入机制的选择问题, 即系统如何根据参数 $\{n, \lambda, D, \beta, L\}$ 选择随机接入机制, 最优网络吞吐性能, 同时控制信吞比。

前面的分析已经指出, 在 PBRA 方案中, 当 $0 < n\lambda < e^{-1}$ 时, 网络处于不饱和情况, 当 $e^{-1} \leq n\lambda$ 时, 网络处于饱和情况; 在 CBRA 方案中, 当 $n\lambda < D$ 时, 网络处于不饱和情况, 当 $D \leq n\lambda$ 时, 网络处于饱和情况, 且有 $1 \leq D$ 。因此, 按参数情况, 以下的讨论分为 3 种情形: $0 < n\lambda < e^{-1}$ 、 $e^{-1} \leq n\lambda < D$ 和 $D \leq n\lambda$ 。

1) 当 $0 < n\lambda < e^{-1}$ 时, 在 PBRA 和 CBRA 方案中, 网络均处于不饱和情况。在这种情况下, 不同方案下吞吐量和信吞比随 $\hat{\lambda}$ 的变化情况如图 5 所示, 图 5 中总输入速率的单位为数据包数/时隙, 吞吐量的单位为比特/时隙, 信吞比的单位为信令数/比特。从图 5(a)可以看出, 两种方案的网络吞吐量相同, 均等于输入速率。从图 5(b)可以看出, 两种方案的信令吞吐量比都在 β 的下方, 说明均满足信吞比约束条件, 但 PBRA 方案中的信吞比更小。因此, 系统应当选择 PBRA 方案。

2) 当 $e^{-1} \leq n\lambda < D$ 时, 在 CBRA 方案中, 网络处于不饱和情况, 而在 PBRA 方案中, 网络处于饱和情况。第 2 节的分析已经指出, 基于 PBRA 的网络可能运作在区域 S_3 或者 S_4 , 但无论如何, 其吞吐量均低于或者等于 e^{-1} 。不同方案下吞吐量 $\hat{\gamma}_{\text{out}}$ 和信吞比 $\frac{S}{L\hat{\lambda}_{\text{out}}}$ 随 $\hat{\lambda}$ 的变化情况如图 6 所

示, 基于 PBRA 的网络的吞吐量低于基于 CBRA 网络的吞吐量, 而且差距随 $n\lambda$ 的增加而增加。另一方面, 得益于持续增长的吞吐量, 如图 6(b)所示, 基于 CBRA 网络的信吞比还随 $n\lambda$ 的增加而降低。因此, 在这种情况下, 5G 系统应当选择 CBRA 方案。

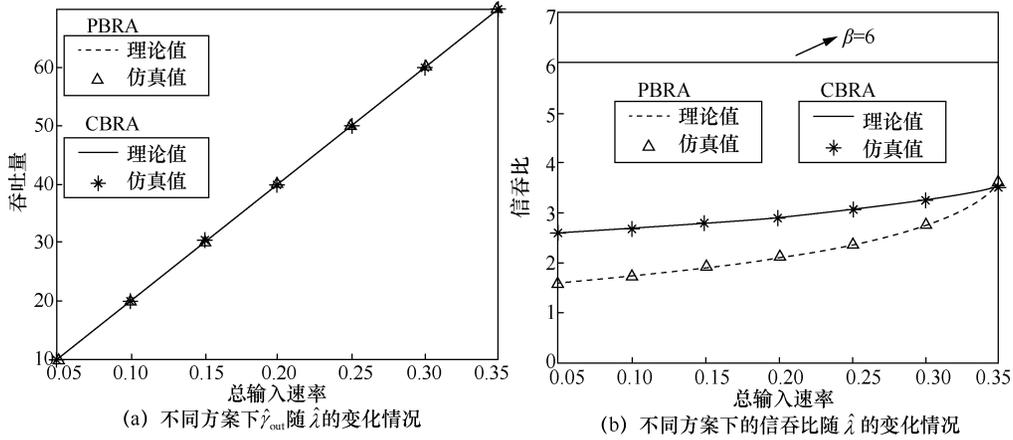


图5 不同方案下吞吐量和信吞比随 $\hat{\lambda}$ 的变化情况 ($n=100$ 、 $s=100$ 、 $\beta=6$ 、 $L=200$ 、CBRA 方案中 $q=0.05$ 、PBRA 方案中 $q=q^*$)

3) 当 $D \leq n\lambda$ 时, 基于 PBRA 的网络或者基于 CBRA 的网络均处于饱和情况。然而, 在 CBRA 方案中, 接入成功的 MTD 会一直占据数据信道传输数据, 导致其他 MTD 无法接入的公平性问题, 不同方案下接入吞吐量 $\hat{\lambda}_{out}^a$ 和信吞比 $\frac{S}{L\hat{\lambda}_{out}}$ 随 $\hat{\lambda}$ 的变化情况如图 7 所示, 图 7 中的接入吞吐量单位为接入请求数/时隙, 从图 7(a)可以看出, 基于 CBRA 的网络的接入吞吐量几乎为零。另一方面, 基于 PBRA 的网络可以通过合理地调整 ACB 因子使得吞吐量处于较高水平, 并且满足信吞比的服务质量要求。因此, 系统此时应当选择 PBRA 方案。

综合以上讨论以及表 1, 本文整理得出基于参数 $\{n, \lambda, D, \beta, L\}$ 的随机接入机制选择与参数配置方案见表 2, 以清晰地阐述系统应该如何根据参数

$\{n, \lambda, D, \beta, L\}$ 选择随机接入机制, 最优化网络吞吐性能, 同时控制信吞比。

5 实际 5G 系统中的应用

从上述理论分析与仿真结果可以看出, 本文所提出的动态随机接入机制选择与参数配置方案可以切实提升 5G 网络对 mMTC 业务的支撑能力。本节将进一步探讨本方案应用到实际 5G 网络的途径。

为了给 mMTC 业务配置合适的参数与随机接入机制, gNB 需要获取 MTD 数目 n 、数据到达率 λ 、数据传输速率 D 、数据报文长度 L 和信吞比上限 β 。在单小区场景中, 所有的 MTD 都与一个 gNB 通信。随机接入机制选择与参数配置方案在 5G 系统中的实施流程如图 8 所示, MTD 首先可以根据 mMTC 业务特征设定信吞比上限 β 。例如, 对于能量受限的设备或者要求低时延的业务, 则上

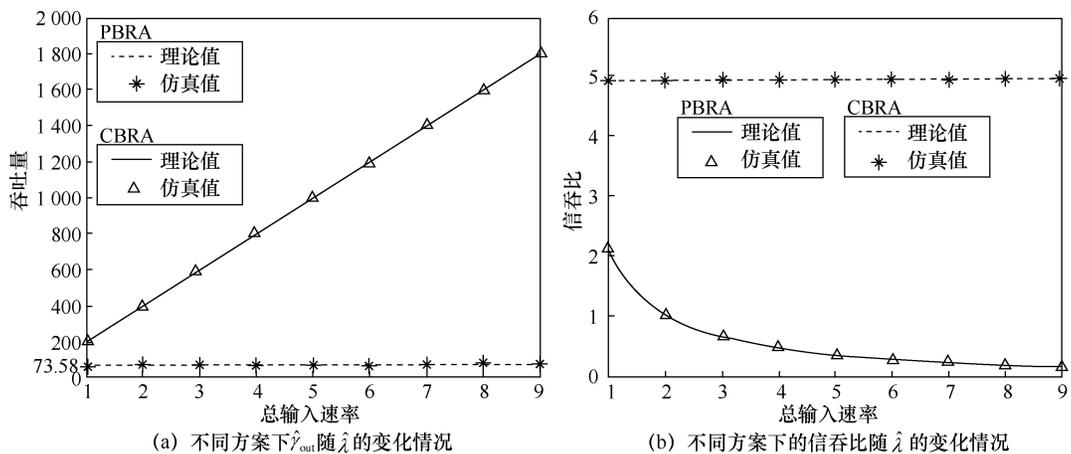


图6 不同方案下吞吐量 $\hat{\gamma}_{out}$ 和信吞比 $\frac{S}{L\hat{\lambda}_{out}}$ 随 $\hat{\lambda}$ 的变化情况 ($n=100$ 、 $s=100$ 、 $L=200$ 、 $q=q^*$ 、 $\beta=6$)

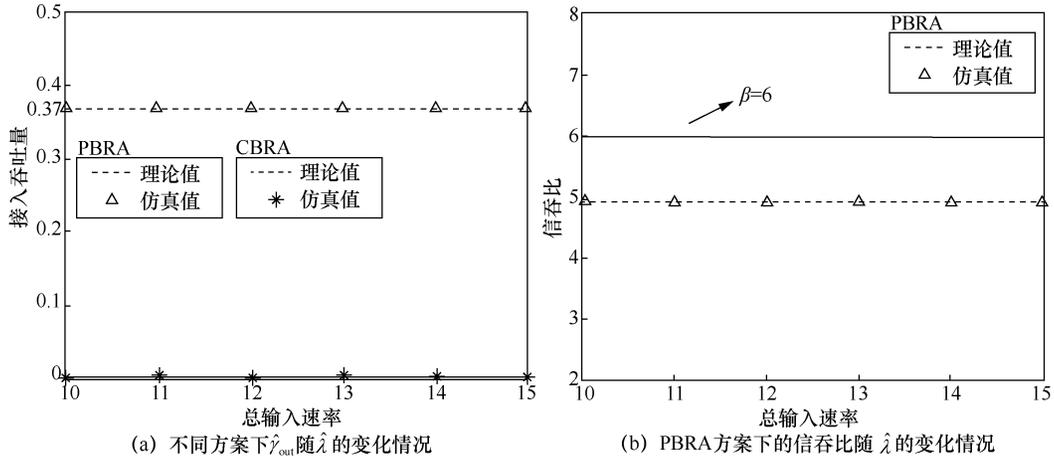


图 7 不同方案下接入吞吐量 $\hat{\lambda}_{out}^a$ 和信吞比 $\frac{S}{L\hat{\lambda}_{out}}$ 随 $\hat{\lambda}$ 的变化情况 ($n=100$ 、 $s=100$ 、 $L=200$ 、 $q=q^*$ 、 $\beta=6$)

表 2 基于参数 $\{n, \lambda, D, \beta, L\}$ 的随机接入机制选择与参数配置方案

$\{n, \lambda, D\}$	$\{\beta, L\}$	q^*	$\hat{\gamma}_{max}$	选择方案
$0 < n\lambda < \frac{1}{e}$	$\frac{4s \exp(-\mathbb{W}_0(-\hat{\lambda})) - s}{L} \leq \beta$	$\left[\frac{\hat{\lambda}}{np_l}, \frac{1}{n} \ln p_s \right]$	$L\hat{\lambda}$	PBRA 方案
	$\frac{4s \exp(-\mathbb{W}_0(-\hat{\lambda})) - s}{L} > \beta$	无解	无解	
$\frac{1}{e} \leq n\lambda < D$	$\frac{4sn\rho q^*(1-p_l) + 5s\hat{\lambda}_{max}^a}{L\hat{\lambda}} \leq \beta$	$q^* = \frac{\hat{\lambda}(D-e^{-1})}{nD(\hat{\lambda}-e^{-1})}$	$L\hat{\lambda}$	CBRA 方案
	$\frac{4sn\rho q^*(1-p_l) + 5s\hat{\lambda}_{max}^a}{L\hat{\lambda}} > \beta$	无解	无解	
$D \leq n\lambda$	$\frac{(4e-1)s}{L} \leq \beta$	$\frac{1}{n}$	Le^{-1}	PBRA 方案
	$\frac{3s}{L} < \beta < \frac{(4e-1)s}{L}$	$\frac{1}{n} \ln \frac{\beta L + s}{4s}$	$\frac{4Ls}{\beta L + s} \ln \frac{\beta L + s}{4s}$	
	$0 < \beta \leq \frac{3s}{L}$	无解	无解	

限值应当小一点。在首次随机接入过程中，每个 MTD 都可以通过 RRC 消息（如随机接入过程中的 RRC 连接建立请求）向 gNB 报告其输入速率 λ 、数据报文长度 L 和信吞比上限 β 。gNB 可以通过 MTD 的 MAC 地址记录已注册的 MTD，从而得到网络中 MTD 数量 n 。数据传输速率 D 则由用户和基站根据信道的服务质量和资源分配等因素协商决定。在获得 $\{n, \lambda, D, \beta, L\}$ 后，gNB 可以基于表 2 中的总结得到最优的随机接入方案和对应的网络参数设置。最后，gNB 可以通过广播的方式告知 MTD 配置信息。

6 结束语

本文分析了 5G 网络中 PBRA 和 CBRA 方案的

吞吐量和信令开销，并以此为基础，解决了面向 mMTC 流量特征的随机接入机制选择问题和参数配置问题。首先，本文推导了在 PBRA 机制下，网络在饱和/不饱和情况下的吞吐量和信吞比的表达式，进一步考虑信吞比的约束，通过解决优化问题，获得了最大吞吐量和最佳的 ACB 因子。仿真结果表明，为了保证信吞比低于上限 β 限制的同时，获得最优的吞吐量性能，网络需要增加数据报文长度，并且需要根据网络参数动态调制 ACB 因子。其次，本文进一步扩展分析到 CBRA 机制的场景，并刻画了网络的吞吐量和信吞比性能。基于上述结果，本文解决了面向 mMTC 业务流量特征的随机接入机制选择问题。仿真结果表明，在总数据输入速率比较小时，应当选择 PBRA 方案；随着总输入

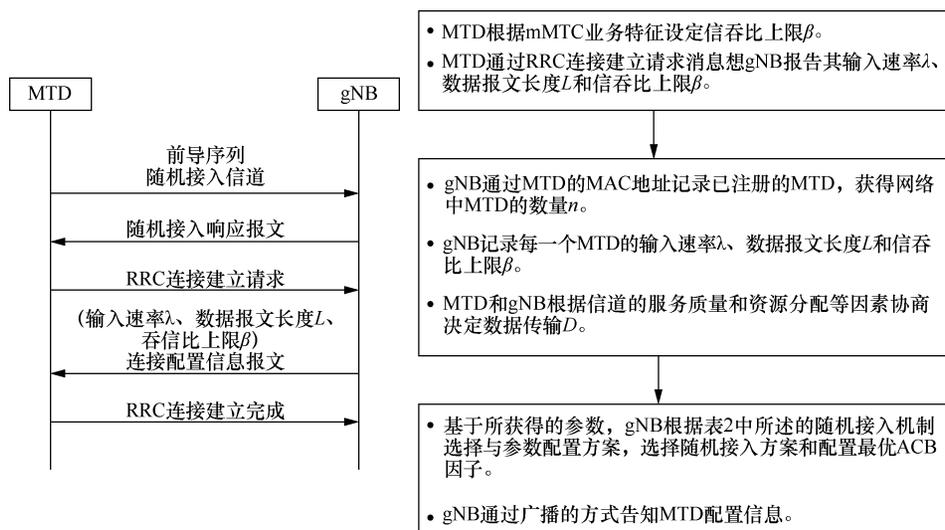


图8 随机接入机制选择与参数配置方案在5G系统中的实施流程

速率的增大, 选择 CBRA 方案有更好的吞吐量和吞信比性能; 当网络总数据输入速率大于数据传输速率时, 网络应该选择 PBRA 方案。本文的分析和仿真结果对需要支持 mMTC 大规模访问的 5G 网络的设计具有重要的意义。

参考文献:

[1] MISIC V B, MISIC J. Machine-to-machine communications[M]. CRC Boca Raton: Press, 2014.

[2] SUMMARYE. Cisco visual networking index: global mobile data traffic forecast update, 2010 - 2015[J]. Growth Lakeland, 2011, 20(4): 2010-2015.

[3] 3GPP. Evolved universal terrestrial radio access (E-UTRA); medium access control (MAC) protocol specification: TS 36.321 V12.5.0[S]. 2015.

[4] LAYA A, ALONSO L, ALONSO-ZARATE J. Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives[J]. IEEE Communications Surveys & Tutorials, 2014, 16(1): 4-16.

[5] KAHN C, VISWANATHAN H. Connectionless access for mobile cellular networks[J]. IEEE Communications Magazine, 2015, 53(9): 26-31.

[6] HOGLUND A, VAN D P, TIRRONEN T, et al. 3GPP release 15 early data transmission[J]. IEEE Communications Standards Magazine, 2018, 2(2): 90-96.

[7] RYOO S, JUNG J, AHN R. Energy efficiency enhancement with RRC connection control for 5G new RAT[C]//Proceedings of 2018 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2018: 1-6.

[8] HAILU S, SAILY M, TIRKKONEN O. RRC state handling for 5G[J]. IEEE Communications Magazine, 2019, 57(1): 106-113.

[9] KHLASS A, LASELVA D, JARVELA R. On the flexible and performance-enhanced radio resource control for 5G NR networks[C]//Proceedings of 2019 IEEE 90th Vehicular Technology Conference. Piscataway: IEEE Press, 2019: 1-6.

[10] FERGUSON M. On the control, stability, and waiting time in a slotted ALOHA random-access system[J]. IEEE Transactions on Communications, 1975, 23(11): 1306-1311.

[11] WU H S, ZHU C X, LA R J, et al. FASA: accelerated S-ALOHA using access history for event-driven M2M communications[J]. IEEE/ACM Transactions on Networking, 2013, 21(6): 1904-1917.

[12] DI C, ZHANG B, LIANG Q L, et al. Learning automata-based access class barring scheme for massive random access in machine-to-machine communications[J]. IEEE Internet of Things Journal, 2019, 6(4): 6007-6017.

[13] LIEN S Y, LIAU T H, KAO C Y, et al. Cooperative access class barring for machine-to-machine communications[J]. IEEE Transactions on Wireless Communications, 2012, 11(1): 27-32.

[14] WANG Z H, WONG V W S. Optimal access class barring for stationary machine type communication devices with timing advance information[J]. IEEE Transactions on Wireless Communications, 2015, 14(10): 5374-5387.

[15] DUAN S Y, SHAH-MANSOURI V, WANG Z H, et al. D-ACB: adaptive congestion control algorithm for bursty M2M traffic in LTE networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(12): 9847-9861.

[16] OH C Y, HWANG D, LEE T J. Joint access control and resource allocation for concurrent and massive access of M2M devices[J]. IEEE Transactions on Wireless Communications, 2015, 14(8): 4182-4192.

[17] ZHAN W, DAI L. Access delay optimization of M2M communications in LTE networks[J]. IEEE Wireless Communications Letters, 2019, 8(6): 1675-1678.

[18] ZHAN W, DAI L. Massive random access of machine-to-machine

- communications in LTE networks: modeling and throughput optimization[J]. IEEE Transactions on Wireless Communications, 2018, 17(4): 2771-2785.
- [19] WAN T, SHEIKH A U. Performance and stability analysis of buffered slotted ALOHA protocols using tagged user approach[J]. IEEE Transactions on Vehicular Technology, 2000, 49(2): 582-593.
- [20] DAI L. Stability and delay analysis of buffered aloha networks[J]. IEEE Transactions on Wireless Communications, 2012, 11(8): 2707-2719.
- [21] ZHAN W, XU C, SUN X H, et al. Toward optimal connection management for massive machine-type communications in 5G system[J]. IEEE Internet of Things Journal, 2021, 8(17): 13237-13250.
- [22] 朱振宇, 朱晓荣, 蔡艳, 等. 基于时隙 ALOHA 与自适应接入类禁止混合的大规模终端接入算法[J]. 物联网学报, 2021, 5(1): 90-98.
- ZHU Z Y, ZHU X R, CAI Y, et al. Large-scale terminal access algorithm based on slot ALOHA and adaptive access class barring[J]. Chinese Journal on Internet of Things, 2021, 5(1): 90-98.
- [23] 3GPP. 5G;NR;Radio resource control (RRC);Protocol specification: TS 38.331 V16.6.0[S]. 2021.
- [24] ZHAN W, DAI L. Massive random access of machine-to-machine communications in LTE networks: throughput optimization with a finite data transmission rate[J]. IEEE Transactions on Wireless Communications, 2019, 18(12): 5749-5763.
- [25] 田乃硕, 徐秀丽, 马占友. 离散时间排队论[M]. 北京: 科学出版社, 2008.
- TIAN N S, XU X L, MA Z Y. Discrete time queuing theory[M]. Beijing: Science Press, 2008.
- [26] LIANG Y W, ZHAN W, SUN X H, et al. Signaling overhead-constrained throughput optimization for 5G packet-based random access with mMTC[C]//Proceedings of 2021 IEEE Global Communications Conference. Piscataway: IEEE Press, 2021: 1-6.

[作者简介]



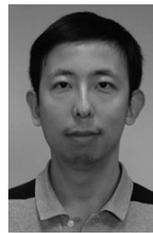
詹文 (1990-), 男, 博士, 中山大学电子与通信工程学院助理教授, 主要研究方向为 5G/6G 网络 and 大规模物联网通信。



梁译文 (1996-), 男, 中山大学电子与通信工程学院硕士生, 主要研究方向为 5G 网络大规模机器类通信的随机接入。



孙兴华 (1985-), 男, 博士, 中山大学电子与通信工程学院副教授, 主要研究方向为下一代无线通信网络、智能通信等。



谭雪松 (1976-), 男, 博士, 电子科技大学通信抗干扰技术国家级重点实验室教授, 主要研究方向包括认知无线网络、抗干扰盲汇聚通信和网络编码在通信中的应用等。