

面向缓存网络内容分发的非正交传输策略研究

杨龙, 李函洋, 任萌萌, 贺冰涛, 周雨晨, 陈健

(西安电子科技大学通信工程学院, 陕西 西安 710071)

摘要: 面向无线缓存网络的内容分发场景, 设计了一种两阶段自适应非正交/正交多址接入传输方案。针对该方案, 提出了一种高可靠性用户调度策略, 并通过理论分析得出了中断概率的解析表达式及其高信噪比下渐近表达式。分析结果表明, 该策略可获得满空间分集增益。仿真结果表明, 与基于瞬时信道状态信息的调度策略、随机调度策略以及传统的正交多址接入传输策略相比, 所提策略可通过获取空间分集增益, 实现优于现有策略的内容分发可靠性。

关键词: 非正交多址接入; 无线缓存; 内容分发; 中断概率; 分集增益

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023193

Research on non-orthogonal transmission strategy for content delivery in caching networks

YANG Long, LI Hanyang, REN Mengmeng, HE Bingtao, ZHOU Yuchen, CHEN Jian

School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

Abstract: A two-stage adaptive non-orthogonal/orthogonal multiple access transmission scheme was designed for content delivery scenarios in wireless caching networks. For the scheme, a high-reliability user scheduling strategy was proposed, and the analytical expression of the outage probability and its asymptotic expression in high signal-to-noise ratio regime were derived by theoretical analysis. The analysis results indicate that the proposed strategy can achieve full spatial diversity gain. Simulation results demonstrate that compared with the instantaneous channel state information-based scheduling strategy, the random scheduling strategy, and the conventional orthogonal multiple access transmission strategy, the proposed strategy can achieve better content delivery reliability than existing strategies by obtaining the spatial diversity gain.

Keywords: non-orthogonal multiple access, wireless caching, content delivery, outage probability, diversity gain

0 引言

无线缓存作为未来无线网络中的一项重要技术^[1], 通过将本地缓存的思路引入传统的无线网络中, 用户请求概率高的热点内容被预先缓存到近用

户端的辅助缓存设备或用户设备自身的本地缓存中, 在内容分发阶段用户可以通过附近的辅助缓存设备以及相邻用户设备间的协作通信获取服务, 避免过多地通过回程链路和核心网络进行通信, 缓解了核心网络的压力, 同时减轻了移动通信流量^[2-4]。

收稿日期: 2023-06-08; 修回日期: 2023-08-24

通信作者: 任萌萌, renmengmeng@stu.xidian.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61971320, No.62201421, No.62371354, No.62371367); 陕西省重点产业创新链基金资助项目 (No.2023-ZDLGY-50); 陕西省自然科学基金基础研究计划基金资助项目 (No.2021JQ-206); 广东省基础与应用基础研究基金资助项目 (No.2020A1515110084)

Foundation Items: The National Natural Science Foundation of China (No.61971320, No.62201421, No.62371354, No.62371367), The Shaanxi Key Industrial Innovation Chain Project (No.2023-ZDLGY-50), The Natural Science Basic Research Plan of Shaanxi Province (No.2021JQ-206), The Guangdong Basic and Applied Basic Research Foundation (No.2020A1515110084)

尽管如此，由于内容分发依赖于无线接入网，而现有移动通信无线接入网均采用正交多址接入 (OMA, orthogonal multiple access)，必须通过资源调度保持用户间资源的时域、频域、码域正交性，难以同时服务大量用户，进而不可避免地产生物理层传输时延。因此，如何优化物理层传输策略，是提升无线缓存内容分发性能的关键。

作为未来无线网络中的另一项关键技术，非正交多址接入 (NOMA, non-orthogonal multiple access) 利用功率域多址思想，通过发送端重叠编码和接收端串行干扰消除 (CIC, cache-enabled interference cancellation)，将不同功率的多信息流叠加传输^[5-7]。与传统 OMA 技术相比，NOMA 技术可在提升频谱效率的同时，支撑更多的用户连接。根据上述原理，若将 NOMA 技术引入缓存内容分发，可在相同时频资源上同时向多用户分发内容，从而显著降低内容分发的物理层传输时延。

目前，已有部分研究工作探索了 NOMA 技术在缓存内容分发方面的应用，具体可分为以下两类。

1) 无基础设施下 NOMA 缓存内容分发。在 NOMA 叠加内容分发时，用户利用本地预先缓存的部分内容进行缓存干扰消除 (SIC, successive interference cancellation)，消除叠加信息中的已缓存部分，从而降低多用户信息流间干扰及信息解析复杂度^[8-9]。文献[8]假设用户已经预先在内容缓存阶段缓存了部分内容，重点分析了在 4 种缓存状态下内容分发阶段的 NOMA 传输方案。文献[9]考虑了具体的车联网场景，提出了一种使用 NOMA 缓存满足车联网中频谱效率需求的方案，同时考虑了完整内容缓存与文件分割缓存场景。另一方面，用户也可利用本地预先缓存内容通过终端直通 (D2D, device-to-device) 互相协助^[10-11]。文献[10]考虑了一个两用户 NOMA 系统，NOMA 能够令每个用户在其他用户的上行链路信道内收到请求的内容，并推导了所提系统的和速率。文献[11]在文献[10]的基础上，提出了系统内容分发总速率最大化的功率分配方案。

2) 有基础设施下 NOMA 缓存内容分发。系统中存在预缓存部分内容的基础设施，如果缓存命中，则直接向服务覆盖范围内的用户提供服务；如果缓存未命中，基站从核心网获取内容后，将内容直接分发给用户。文献[12-13]分析了多用户的 NOMA 缓存内容分发场景。文献[12]假设允许重用

分发完成的内容文件的未占用功率，设计了一种基于深度学习的方法使分发时延最小化并提升计算效率。文献[13]考虑的是无线缓存网络的下行传输场景，分析了典型用户的覆盖概率，并通过仿真表明了 NOMA 为无线缓存网络带来了性能增益。文献[14-15]主要针对两用户情形。文献[14]提出了一个面向服务质量 (QoS, quality of service) 的动态功率分配策略，在内容分发阶段使用 NOMA 同时为 2 个用户提供服务，功率分配的目标是确保远端用户的 QoS 目标。文献[15]提出了 2 种 NOMA 辅助缓存的策略，即先推送后分发与推送同时分发。第一种策略下缓存服务器通过 NOMA 在某一时刻内对 2 个用户完成内容分发；第二种策略下基站在为某个用户直接服务的同时通过 NOMA 对缓存服务器进行内容推送。

在上述有基础设施下 NOMA 缓存内容分发研究中，文献[12-14]未考虑缓存未命中用户内容分发。然而，在实际的无线信道传输中，由于信号传输受到发送功率以及实际信道中的衰落的影响，推送的所有内容未必全部能被成功缓存。文献[15]尽管考虑了缓存未命中情况，但其所提策略下缓存未命中用户将在远距离基站发送的 NOMA 叠加信号中解码所需内容，意味着该用户不仅面临信道条件差的问题，还要与其他缓存服务器共享资源块。在实际中，如果通过 NOMA 向多个设备分发内容，用户接收机处可能会存在误差传播、解码复杂度高以及信号处理时延高等问题。如果对用户进行调度，则可以最大限度缓解上述问题带来的影响。现有用户调度策略主要面向非缓存的无线通信场景^[16]，如何通过用户调度并利用信道状态多样性实现空间分集从而提升无线缓存场景下内容分发的传输可靠性，目前仍未有研究。

针对现有研究的局限性，本文面向有基础设施 NOMA 无线缓存的内容分发阶段，研究可靠传输以及用户调度问题，贡献如下。

1) 自适应传输方案设计。考虑缓存命中用户与缓存未命中用户共存场景，提出了一种两阶段自适应 NOMA/OMA 协作传输方案，由缓存服务器充当协作节点，在服务缓存命中用户的同时自适应地向缓存未命中用户转发内容信息。

2) 高可靠用户调度策略实现。针对所提传输方案的 NOMA 与 OMA 情况，分别提出了对应的高可靠性用户调度准则，并给出了所提调度策略最优性证明过程以及具体的用户调度实现策略。

3) 分发可靠性分析。分别进行了中断概率分析

以及渐近分析,证明了根据所提调度策略确定的调度用户在高信噪比下能够实现满分集增益;由于缓存未命中用户在解码过程中使用了最大比合并(MRC, maximal ratio combining),因此在高信噪比下可以获取2阶分集增益。

4) 仿真验证及性能对比。仿真结果验证了本文中中断概率以及分集增益推导的正确性。此外,通过与基于瞬时信道状态信息(CSI, channel state information)的调度策略、OMA传输策略以及随机调度策略进行对比,证明了本文所提传输方案与用户调度策略在中断性能上的优越性。

1 系统模型

考虑无线缓存网络内容分发场景,内容分发阶段系统模型如图1所示。该系统具有一个基站BS、一个缓存服务器CS以及 $K+1$ 个待服务用户,其中,用户可分为 K 个缓存命中用户与一个缓存未命中用户。缓存命中用户请求的内容在内容推送阶段已被缓存到CS中,在内容分发阶段可由该缓存服务器直接提供服务。用 $\mathcal{K} \triangleq \{1, \dots, K\}$ 表示 K 个缓存命中用户的集合, $u_k (k \in \mathcal{K})$ 表示缓存命中用户集合中的第 k 个用户。缓存未命中用户无法在缓存服务器CS中直接找到请求的文件,该用户用 u_0 表示。假设BS与CS间链路的信道系数为 f ,BS与用户 u_0, u_1, \dots, u_K 间链路的信道系数分别是 h_0, h_1, \dots, h_K ,CS与用户 u_0, u_1, \dots, u_K 间链路的信道系数分别是 g_0, g_1, \dots, g_K ,每一种链路都服从独立、同分布的准静态瑞利块衰落,不同链路的分布参数不同^[17]。具体地,信道增益 $|f|^2$ 服从均值为 F 的指数分布, $|h_0|^2, |h_1|^2, \dots, |h_K|^2$ 分别服从均值为 H_0, H_1, \dots, H_K 的指数分布, $|g_0|^2, |g_1|^2, \dots, |g_K|^2$ 分别服从均值为 G_0, G_1, \dots, G_K 的指数分布。

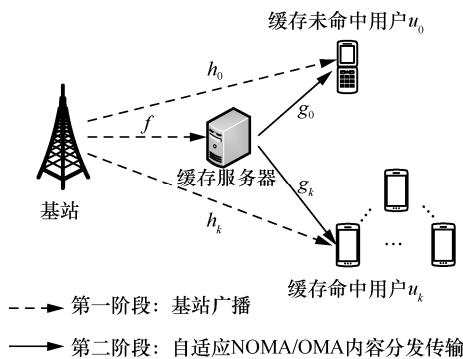


图1 内容分发阶段系统模型

2 两阶段自适应 NOMA/OMA 传输方案

假设总传输时延为 T ,记 $x_k (k \in \mathcal{K})$ 为缓存命中用户 $u_k (k \in \mathcal{K})$ 请求的内容,内容信息量记为 $V_k (k \in \mathcal{K})$,则对应的目标速率为 $r_k = \frac{V_k}{T, (k \in \mathcal{K})}$,解码门限为 $\tau_k = 2^{r_k} - 1, k \in \mathcal{K}$ 。同样地,记 x_0 为缓存未命中用户 u_0 请求的内容,内容信息量记为 V_0 ,则对应的目标速率为 $r_0 = \frac{V_0}{T}$,解码门限为 $\tau_0 = 2^{r_0} - 1$ 。

2.1 第一阶段：基站广播

在第一阶段,基站BS广播内容 x_0 ,基站BS的广播信号记为 $\sqrt{P_B}x_0$,其中 P_B 为BS的发送功率。则缓存服务器CS的接收信号可表示为 $y_{CS} = \sqrt{P_B}fx_0 + n_{CS}$,其中, n_{CS} 表示在缓存服务器CS处的加性白高斯噪声,其方差为 σ^2 。缓存服务器CS的接收信噪比为 $\gamma_{BS \rightarrow CS,0} = \frac{P_B |f|^2}{\sigma^2}$ 。用户 u_0 的接收信号可以表示为 $y_{u_0} = \sqrt{P_B}h_0x_0 + n_{u_0}$,其中, n_{u_0} 表示在用户 u_0 处的加性白高斯噪声,其方差为 σ^2 ,基于此可得到用户 u_0 的接收信噪比为 $\gamma_{BS \rightarrow u_0,0} = \frac{P_B |h_0|^2}{\sigma^2}$ 。用户 u_k 的接收信号可以表示为 $y_{u_k} = \sqrt{P_B}h_kx_0 + n_{u_k}$,其中, n_{u_k} 表示在用户 u_k 处的加性白高斯噪声,其方差为 σ^2 ,基于此可得到用户 u_k 的接收信噪比为 $\gamma_{BS \rightarrow u_k,0} = \frac{P_B |h_k|^2}{\sigma^2}$ 。

2.2 第二阶段：自适应 NOMA/OMA 内容分发传输

在第二阶段,CS根据第一阶段信号的收发情况从缓存命中用户中调度一个用户,记为 u_k ;然后,CS将自适应选择以NOMA或OMA方式发送用户 u_0 和 u_k 期望的内容。其他未调度的缓存命中用户可能会在下一次传输过程中被机会式调度。具体分析如下。

1) CIC和SIC的NOMA内容分发

在第一阶段中,当CS成功解码出内容 x_0 ,即 $\gamma_{BS \rightarrow CS,0} \geq \tau_0$ 时,在本阶段CS将发送一个 x_k 与 x_0 的NOMA叠加信号到用户 u_k 和用户 u_0 ,该信号记为 $\sqrt{P_C \alpha}x_k + \sqrt{P_C(1-\alpha)}x_0$,其中, P_C 为CS的发送功率, α 为分配给内容 x_k 的功率分配系数。由于选出的调度用户 u_k 需要与缓存未命中用户 u_0 在同一传输资

源块下获取服务，为了保证调度用户 u_k 的接收可靠性，CS 将大部分功率分配给 x_k 。

对于用户 u_k ，其接收信号可以表示为 $y'_{u_k} = \sqrt{P_C \alpha} g_k x_k + \sqrt{P_C(1-\alpha)} g_0 x_0 + n_{u_k}$ ；然后，根据第一阶段是否成功解码内容 x_0 ，用户 u_k 将采取不同的解码方式。具体为，若 u_k 在第一阶段未成功解码 x_0 ，则在本阶段通过正常的 SIC 进行解码，即 u_k 将发送给 u_0 的信号 x_0 作为干扰，直接解码自己的内容 x_k ；否则，用户 u_k 在本阶段解码时可通过 CIC 技术利用前一阶段的缓存直接去除内容 x_0 带来的干扰。因此，用户 u_k 解码 x_k 的信干噪比可以表示为

$$\gamma_{CS \rightarrow u_k, k} = \begin{cases} \frac{P_C |g_k|^2 \alpha}{P_C |g_k|^2 (1-\alpha) + \sigma^2}, & \gamma_{BS \rightarrow u_k, 0} < \tau_0 \\ \frac{P_C |g_k|^2 \alpha}{\sigma^2}, & \gamma_{BS \rightarrow u_k, 0} \geq \tau_0 \end{cases} \quad (1)$$

对于用户 u_0 ，在接收到 NOMA 叠加信号后首先将内容 x_0 作为干扰解码内容 x_k ，对应的解码信干噪比为 $\gamma_{CS \rightarrow u_0, k} = \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2}$ 。如果 x_k 被成功解码，则 u_0 将其从干扰中移除并继续解码期望内容 x_0 ，此时解码信干噪比可以表示为 $\gamma_{CS \rightarrow u_0, 0} = \frac{P_C |g_0|^2 \alpha}{\sigma^2}$ 。如果 x_k 未被成功解码，则缓存未命中用户 u_0 的解码过程中断。结合 x_k 的解码情况，用户 u_0 解码 x_0 的信干噪比可以表示为

$$\gamma_{CS \rightarrow u_0, 0} = \begin{cases} \frac{P_C |g_0|^2 \alpha}{\sigma^2}, & \gamma_{CS \rightarrow u_0, k} \geq \tau_k \\ 0, & \gamma_{CS \rightarrow u_0, k} < \tau_k \end{cases} \quad (2)$$

2) OMA 内容分发

在第一阶段，当 CS 未能成功解码内容 x_0 ，即 $\gamma_{BS \rightarrow CS, 0} < \tau_0$ 时，在本阶段缓存服务器 CS 将通过 OMA 的方式发送内容 x_k 到用户 u_k ，发送信号可记为 $\sqrt{P_C} x_k$ 。

此时对于用户 u_k 来说，解码 x_k 的信干噪比表示为 $\gamma'_{CS \rightarrow u_k, k} = \frac{P_C |g_k|^2}{\sigma^2}$ 。对于用户 u_0 ，在本阶段采用 OMA 传输的情况下并没有收到内容 x_0 ，因此没有解码 x_0 的过程。

2.3 两阶段自适应传输结果

对于用户 u_0 ，采用 MRC 计算两阶段解码 x_0 的

总信干噪比；对于用户 u_k ，则只需要考虑第二阶段解码 x_k 的信干噪比。

在第二阶段采用联合 CIC 和 SIC 的 NOMA 内容分发（条件为 $\gamma_{BS \rightarrow CS, 0} \geq \tau_0$ ）时，可以分别得到用户 u_0 和用户 u_k 在 NOMA 阶段的解码信干噪比为

$$\gamma_0^{\text{NOMA}} = \gamma_{BS \rightarrow u_0, 0} + \gamma_{CS \rightarrow u_0, 0} = \begin{cases} \frac{P_B |h_0|^2}{\sigma^2} + \frac{P_C |g_0|^2 \alpha}{\sigma^2}, & \gamma_{CS \rightarrow u_0, k} \geq \tau_k \\ \frac{P_B |h_0|^2}{\sigma^2}, & \gamma_{CS \rightarrow u_0, k} < \tau_k \end{cases} \quad (3)$$

$$\gamma_k^{\text{NOMA}} = \gamma_{CS \rightarrow u_k, k} = \begin{cases} \frac{P_C |g_k|^2 \alpha}{P_C |g_k|^2 (1-\alpha) + \sigma^2}, & \gamma_{BS \rightarrow u_k, 0} < \tau_0 \\ \frac{P_C |g_k|^2 \alpha}{\sigma^2}, & \gamma_{BS \rightarrow u_k, 0} \geq \tau_0 \end{cases} \quad (4)$$

在采用 OMA 内容分发（条件为 $\gamma_{BS \rightarrow CS, 0} < \tau_0$ ）时，得到用户 u_0 和 u_k 在 OMA 传输时的解码信噪比为

$$\gamma_0^{\text{OMA}} = \gamma_{BS \rightarrow u_0, 0} = \frac{P_B |h_0|^2}{\sigma^2} \quad (5)$$

$$\gamma_k^{\text{OMA}} = \gamma'_{CS \rightarrow u_k, k} = \frac{P_C |g_k|^2}{\sigma^2} \quad (6)$$

在由 LTE/5G 驱动的实际无线缓存/通信系统中，通常存在多个正交资源块，这些正交资源块将被分配给不同的用户。由于不同正交资源块之间的正交性，不同缓存未命中用户的内容分发过程相互独立，因此本文重点考虑针对一个典型缓存未命中用户的内容分发方案设计。然而，当系统中存在多个缓存未命中用户时，可先将缓存命中用户分为多个无重叠的缓存命中用户组；其次，每个缓存未命中用户将会与一个缓存命中用户组配对。在上述假设下，每个配对组合可以执行本文所设计的两阶段自适应 NOMA/OMA 传输方案及用户调度策略。

3 高可靠性的用户调度策略及最优性证明

上述两阶段自适应 NOMA/OMA 传输方案在调度用户已知的情况下，给出了具体的通信过程，为了最大化用户可靠性，本文设计了面向高可靠性的用户调度策略，该策略针对 NOMA 和 OMA 传输过程具有不同的用户调度准则。具体的调度准则以及最优性证明如下。

针对 NOMA 情况，应采用如准则 1 所示的调度策略。

准则 1

$$k^* = \arg \max_{k \in \mathcal{K}} X_k \quad (7)$$

X_k 的定义如下

$$X_k \triangleq |g_k|^2 \left\{ \frac{\alpha}{\tau_k} - \bar{\alpha} [1 - \varepsilon(J_k)] \right\} \quad (8)$$

其中， $\bar{\alpha} = 1 - \alpha$, $J_k \triangleq P_B |h_k|^2 - \tau_0 \sigma^2$; $\varepsilon(J_k)$ 是一个阶跃函数，当 $J_k \geq 0$ 时 $\varepsilon(J_k) = 1$ ，否则 $\varepsilon(J_k) = 0$ 。

引理 1 在 NOMA 情况下，用户 u_k 的中断概率关于 X_k 单调递减。

证明 由对传输细节的分析可知，NOMA 情况下用户 u_k 的中断事件表示为 $\{\gamma_{BS \rightarrow CS,0} \geq \tau_0, \gamma_k^{\text{NOMA}} < \tau_k\}$ ，将信干噪比的表达式代入中断事件中，可得出 NOMA 情况下用户 u_k 的中断概率表达式为

$$P_{\text{out},u_k}^{\text{NOMA}} = \Pr\{\gamma_{BS \rightarrow CS,0} \geq \tau_0, \gamma_k^{\text{NOMA}} < \tau_k\} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_C |g_k|^2 \alpha}{P_C |g_k|^2 \bar{\alpha} + \sigma^2} < \tau_k, \frac{P_B |h_k|^2}{\sigma^2} < \tau_0\right\} + \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_C |g_k|^2 \alpha}{\sigma^2} < \tau_k, \frac{P_B |h_k|^2}{\sigma^2} \geq \tau_0\right\} \quad (9)$$

其中，触发 NOMA 传输的条件事件为 $\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\}$ ，根据本文对信道状况的假设，BS 与 CS 之间链路的信道增益 $|f|^2$ 以及 CS 与各用户之间链路的信道增益 $|g_k|^2$ 服从相互独立的指数分布，因此式(9)可以写为

$$P_{\text{out},u_k}^{\text{NOMA}} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\} \left(P_{\text{out},u_k}^{\text{NOMA,SIC}} + P_{\text{out},u_k}^{\text{NOMA,CIC}}\right) \quad (10)$$

$$P_{\text{out},u_k}^{\text{NOMA,SIC}} = \Pr\left\{\frac{P_C |g_k|^2 \alpha}{P_C |g_k|^2 (1-\alpha) + \sigma^2} < \tau_k, \frac{P_B |h_k|^2}{\sigma^2} - \tau_0 < 0\right\} \quad (11)$$

$$P_{\text{out},u_k}^{\text{NOMA,CIC}} = \Pr\left\{\frac{P_C |g_k|^2 \alpha}{\sigma^2} < \tau_k, \frac{P_B |h_k|^2}{\sigma^2} - \tau_0 \geq 0\right\} \quad (12)$$

其中， $P_{\text{out},u_k}^{\text{NOMA,SIC}}$ 表示 NOMA 内容推送下用户 u_k 使

用 SIC 解码的中断概率， $P_{\text{out},u_k}^{\text{NOMA,SIC}}$ 表示 NOMA 内容推送下用户 u_k 使用 CIC 解码的中断概率。已知式(11)与式(12)代表的事件互斥，由于 2 个互斥事件的概率之和等于其和事件的概率。因此式(9)可以等价于

$$P_{\text{out},u_k}^{\text{NOMA}} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\} \cdot \Pr\left\{|g_k|^2 \left\{\frac{\alpha}{\tau_k} - \bar{\alpha} [1 - \varepsilon(J_k)]\right\} < \frac{\sigma^2}{P_C}\right\} \quad (13)$$

将式(8)代入式(13)可得

$$P_{\text{out},u_k}^{\text{NOMA}} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\} \Pr\left\{X_k < \frac{\sigma^2}{P_C}\right\} \quad (14)$$

由于 $\frac{\sigma^2}{P_C}$ 是一个常量，因此挑选出的用户 u_k 对应的 X_k 越大，NOMA 情况下该用户的中断概率越小。

针对 OMA 情况，应采用如准则 2 所示的调度策略。

准则 2

$$k^* = \arg \max_{k \in \mathcal{K}} Y_k \quad (15)$$

Y_k 的定义如下

$$Y_k \triangleq \frac{|g_k|^2}{\tau_k} \quad (16)$$

引理 2 在 OMA 情况下，用户 u_0 中断概率关于 Y_k 单调递减。

证明 由之前的分析可知，在 OMA 情况下，缓存服务器将以全功率发送内容 x_k 给用户 u_k ，此时用户 u_k 的中断事件表示为 $\{\gamma_{BS \rightarrow CS,0} < \tau_0, \gamma_k^{\text{OMA}} < \tau_k\}$ ，将对应的信干噪比的表达式代入中断事件中，可得出 OMA 情况下用户 u_k 的中断概率为

$$P_{\text{out},u_k}^{\text{OMA}} = \Pr\{\gamma_{BS \rightarrow CS,0} < \tau_0, \gamma_k^{\text{OMA}} < \tau_k\} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0, \frac{P_C |g_k|^2}{\sigma^2} < \tau_k\right\} \quad (17)$$

其中，触发 OMA 传输的条件事件表示为 $\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0\right\}$ ，由于 $|f|^2$ 与 $|g_k|^2$ 服从相互独立的指数分布，根据独立性，式(17)等价于

$$P_{\text{out},u_k}^{\text{OMA}} = \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} < \tau_0 \right\} \Pr \left\{ \frac{|g_k|^2}{\tau_k} < \frac{\sigma^2}{P_C} \right\} \quad (18)$$

将式(16)代入式(18)可得

$$P_{\text{out},u_k}^{\text{OMA}} = \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} < \tau_0 \right\} \Pr \left\{ Y_k < \frac{\sigma^2}{P_C} \right\} \quad (19)$$

由于 $\frac{\sigma^2}{P_C}$ 是一个常量，因此挑选出的用户 u_k 对应的 Y_k 越大，OMA 情况下该用户的中断概率越小。

4 用户调度具体实现策略

在提出了调度策略并完成了最优性证明之后，本节主要讲述两阶段自适应 NOMA/OMA 传输方案与用户调度的具体实现策略，并详细描述信道状态信息获取的过程。与文献[16]类似，用户调度具体实现策略如图 2 所示，本文中考虑的是一个块衰落信道模型，任何链路的信道增益在一个持续时间 T 内不会改变。第一阶段基站广播持续的时间为 T_1 ，第二阶段自适应 NOMA/OMA 传输持续的时间为 T_2 ，余下部分用于用户调度过程。用户调度过程具体分为 3 个子时隙，即子时隙 1~子时隙 3。

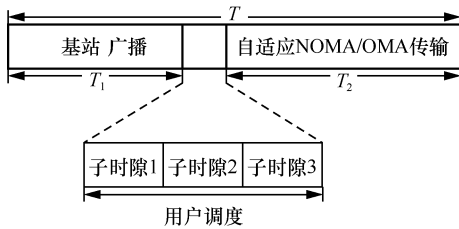


图 2 用户调度具体实现策略

1) 在子时隙 1，基站 BS 广播一个信令消息给所有用户，场景中的所有用户以及缓存服务器 CS 利用该信息进行信道估计，用户获取自身与基站之间链路的信道增益 $|h_0|^2$ 与 $|h_k|^2$ ($k \in \mathcal{K}$)，而缓存服务器 CS 除了获取自身与基站之间链路的信道增益 $|f|^2$ 外还要判断是否可以成功解码内容 x_0 ，以此来确定之后采取 NOMA 传输还是 OMA 传输。

2) 在子时隙 2，缓存服务器 CS 在完成判断后，广播一个信令信息 flag_success 来告知其他用户第二阶段的传输方案。

3) 在子时隙 3，假设该时隙的持续时间为 t 。其他用户在接收到信令信息 flag_success 后，根据不同结果采取对应行动，具体如下。

$\text{flag_success} = 1$ ，即第一阶段缓存服务器 CS 成功缓存了内容 x_0 。第二阶段的传输方案为 NOMA，用户 u_k ($k \in \mathcal{K}$) 根据 X_k 的结果构造一个虚拟定时器，初始化为 $\omega_k = t \exp(-X_k) < t$ ，用户根据定时器反馈信令，最先反馈的用户即依据准则 $k^* = \arg \max X_k$ 调度出来的用户。

$\text{flag_success} = 0$ ，即第一阶段缓存服务器 CS 未成功缓存内容 x_0 。第二阶段的传输方案为 OMA，用户 u_k ($k \in \mathcal{K}$) 根据 Y_k 的结果构造一个虚拟定时器，初始化为 $\omega_k = t \exp(-Y_k) < t$ ，用户根据定时器反馈信令，最先反馈的用户即依据准则 $k^* = \arg \max Y_k$ 调度出来的用户。

当调度用户确定后，该用户广播一个信令信息用以告知缓存服务器 CS 以及其他所有用户。

计算复杂度分析如下。根据用户调度具体实现策略，在子时隙 1，CS 需要计算是否可以成功解码 x_0 ；在子时隙 3，每个缓存命中用户 u_k ($k \in \mathcal{K}$) 将计算 ω_k 。因此，本文所提调度策略的计算复杂度为 $O(K)$ 。

时间复杂度分析如下。在子时隙 1，BS 广播一个信令消息；在子时隙 2，CS 广播信令消息 flag_success ；在子时隙 3，虚拟定时器倒计时最小的缓存命中用户最先反馈信令信息。因此，本文所提调度策略的通信开销为 $O(3)$ 。此外，由于每个信令信息包含的比特数很小，因此其时间复杂度可忽略不计。

5 性能分析

5.1 调度用户的中断概率分析

将通过式(7)或式(15)确定的用户称为调度用户，记作 u_{k^*} 。第一阶段结束后无论进入 NOMA 传输还是 OMA 传输，调度用户 u_{k^*} 都会在第二阶段接收由 CS 发送的包含其请求内容的信号。具体地，在 NOMA 情况下，CS 发送 x_{k^*} 和 x_0 的叠加信号， u_{k^*} 将根据其在第一阶段对内容 x_0 的缓存情况决定使用 CIC 或者 SIC 对内容 x_{k^*} 进行解码；而在 OMA 情况下，CS 只发送内容 x_{k^*} ， u_{k^*} 将对接收信号进行直接解码。综上，调度用户 u_{k^*} 的中断事件为

$$\left\{ \gamma_{\text{BS} \rightarrow \text{CS},0} \geq \tau_0, \gamma_{k^*}^{\text{NOMA}} < \tau_{k^*} \right\} \cup \left\{ \gamma_{\text{BS} \rightarrow \text{CS},0} < \tau_0, \gamma_{k^*}^{\text{OMA}} < \tau_{k^*} \right\}.$$

因此，根据中断事件可知调度用户 u_{k^*} 的中断概率为

$$\begin{aligned}
 P_{\text{out},u_{k^*}} &= \Pr\left\{\gamma_{\text{BS}\rightarrow\text{CS},0} \geq \tau_0, \gamma_{k^*}^{\text{NOMA}} < \tau_{k^*}\right\} + \\
 &\Pr\left\{\gamma_{\text{BS}\rightarrow\text{CS},0} < \tau_0, \gamma_{k^*}^{\text{OMA}} < \tau_{k^*}\right\} = \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_C |g_{k^*}|^2 \alpha}{P_C |g_{k^*}|^2 \bar{\alpha} + \sigma^2} < \tau_{k^*}, \right. \\
 &\quad \left. \frac{P_B |h_{k^*}|^2}{\sigma^2} < \tau_0\right\} + \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_C |g_{k^*}|^2 \alpha}{\sigma^2} < \tau_{k^*}, \right. \\
 &\quad \left. \frac{P_B |h_{k^*}|^2}{\sigma^2} \geq \tau_0\right\} + \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0, \frac{P_C |g_{k^*}|^2}{\sigma^2} < \tau_{k^*}\right\} \quad (20)
 \end{aligned}$$

根据本文提出的调度策略可知 $X_{k^*} = \max_{k \in \mathcal{K}} X_k$,

$Y_{k^*} = \max_{k \in \mathcal{K}} Y_k$, 因此式(20)等价于

$$\begin{aligned}
 P_{\text{out},u_{k^*}} &= \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\} \Pr\left\{\max_{k \in \mathcal{K}} X_k < \frac{\sigma^2}{P_C}\right\} + \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0\right\} \Pr\left\{\max_{k \in \mathcal{K}} Y_k < \frac{\sigma^2}{P_C}\right\} = \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0\right\} \prod_{k \in \mathcal{K}} \Pr\left\{X_k < \frac{\sigma^2}{P_C}\right\} + \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0\right\} \prod_{k \in \mathcal{K}} \Pr\left\{Y_k < \frac{\sigma^2}{P_C}\right\} \quad (21)
 \end{aligned}$$

为了方便计算, 在求解时利用

$\Pr\left\{X_k < \frac{\sigma^2}{P_C}\right\} = P_{\text{out},u_k}^{\text{NOMA,SIC}} + P_{\text{out},u_k}^{\text{NOMA,CIC}}$ 的等价关系进行转化, 已知所有链路的信道增益都服从相互独立的指数分布^[18], 因此可以求出调度用户 u_{k^*} 的中断概率数学闭式解为

$$\begin{aligned}
 P_{\text{out},u_{k^*}} &= e^{-\frac{\tau_0}{\rho F}} \prod_{k=1}^K \left[\left(1 - e^{-\frac{\tau_k}{\mu \rho G_k (\alpha - \bar{\alpha} \tau_k)}} \right) \left(1 - e^{-\frac{\tau_0}{\rho H_k}} \right) + \right. \\
 &\quad \left. \left(1 - e^{-\frac{\tau_k}{\mu \rho G_k \alpha}} \right) e^{-\frac{\tau_0}{\rho H_k}} \right] + \left(1 - e^{-\frac{\tau_0}{\rho F}} \right) \prod_{k=1}^K \left(1 - e^{-\frac{\tau_k}{\mu \rho G_k}} \right) \quad (22)
 \end{aligned}$$

其中, 为了统一表达, 设 $\rho = \frac{P_B}{\sigma^2}$ 表示基站的发送信噪比, 同时设 $\mu = \frac{P_C}{P_B}$, 后文将直接采用这 2 个等价关系。

5.2 缓存未命中用户的中断概率分析

缓存未命中用户 u_0 通过两阶段 MRC 完成对内容 x_0 的解码^[19]。第一阶段中基站将广播内容文件 x_0 , 缓存未命中用户 u_0 可尝试接收。缓存服务器对内容 x_0 的解码情况将决定第二阶段的具体传输方式。当第二阶段采用 NOMA 的联合分发形式时, 缓存服务器 CS 将使用 NOMA 原理对内容 x_k 、 x_0 进行叠加并发送。用户 u_0 优先解码分配较大功率的内容文件 x_k , 如果成功解码, 将其作为干扰消除并尝试接收内容 x_0 , 最终根据两阶段的 MRC 判断缓存未命中用户 u_0 是否发生中断。当第二阶段为 OMA 传输时, 缓存服务器 CS 只发送内容文件 x_k , 最终根据第一阶段对内容 x_0 的接收情况判断缓存未命中用户 u_0 是否发生中断。综上可知, 缓存未命中用户 u_0 的中断事件可以表示为 $\left\{\gamma_{\text{BS}\rightarrow\text{CS},0} \geq \tau_0, \gamma_0^{\text{NOMA}} < \tau_0\right\} \cup \left\{\gamma_{\text{BS}\rightarrow\text{CS},0} < \tau_0, \gamma_0^{\text{OMA}} < \tau_0\right\}$, 根据中断事件可知缓存未命中用户 u_0 的中断概率为

$$P_{\text{out},u_0} = P_{\text{out},u_0}^{\text{NOMA}} + P_{\text{out},u_0}^{\text{OMA}} \quad (23)$$

将 NOMA 传输与 OMA 传输分开讨论, 其中

$$P_{\text{out},u_0}^{\text{OMA}} = \Pr\left\{\frac{P_B |f|^2}{\sigma^2} < \tau_0, \frac{P_B |h_0|^2}{\sigma^2} < \tau_0\right\} \quad (24)$$

$$\begin{aligned}
 P_{\text{out},u_0}^{\text{NOMA}} &= \Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \right. \\
 &\quad \left. \frac{P_B |h_0|^2}{\sigma^2} + \frac{P_C |g_0|^2 \bar{\alpha}}{\sigma^2} < \tau_0, \right. \\
 &\quad \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 \bar{\alpha} + \sigma^2} \geq \tau_{k^*}\right\} + \\
 &\Pr\left\{\frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_B |h_0|^2}{\sigma^2} < \tau_0, \right. \\
 &\quad \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1 - \alpha) + \sigma^2} < \tau_{k^*}\right\} \quad (25)
 \end{aligned}$$

OMA 情况下 u_0 的中断概率的闭式解为

$$P_{\text{out},u_0}^{\text{OMA}} = \left(1 - e^{-\frac{\tau_0}{\rho F}} \right) \left(1 - e^{-\frac{\tau_0}{\rho H_0}} \right) \quad (26)$$

接下来分析 NOMA 情况下 u_0 的中断概率 $P_{\text{out},u_0}^{\text{NOMA}}$ 的闭式解推导过程。与调度策略有关的量为 τ_{k^*} , 将其用概率的形式表示为

$$\begin{aligned}
P_{\text{out},u_0}^{\text{NOMA}} = & \sum_{i=1}^K \left\{ \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \right. \right. \\
& \frac{P_B |h_0|^2}{\sigma^2} + \frac{P_C |g_0|^2 (1-\alpha)}{\sigma^2} < \tau_0, \\
& \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2} \geq \tau_i, k^* = i \right\} + \\
& \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_B |h_0|^2}{\sigma^2} < \tau_0, \right. \\
& \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2} < \tau_i, k^* = i \right\} \quad (27)
\end{aligned}$$

其中，事件 $\{k^* = i\}$ 可以等价于

$$\{k^* = i\} = \left\{ \bigcap_{j \in \{\mathcal{K} \setminus \{i\}\}} X_j < X_i \right\} \quad (28)$$

为了以更加简单的形式表达概率 $P_{\text{out},u_0}^{\text{NOMA}}$ ，令

$$\begin{aligned}
P_i^{N1} = & \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_B |h_0|^2}{\sigma^2} + \frac{P_C |g_0|^2 (1-\alpha)}{\sigma^2} < \tau_0, \right. \\
& \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2} \geq \tau_i \right\}, P_i^{N2} = \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \right. \\
& \left. \frac{P_B |h_0|^2}{\sigma^2} < \tau_0, \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2} < \tau_i \right\}, \quad \text{根据} \\
& X_k (k \in \mathcal{K}) \text{ 的定义可知事件 } \{k^* = i\} \text{ 与式(27)中的其他事件相互独立, 综上, 式(27)可以等价于}
\end{aligned}$$

$$P_{\text{out},u_0}^{\text{NOMA}} = \sum_{i=1}^K \left\{ P_i^{N1} \Pr\{k^* = i\} + P_i^{N2} \Pr\{k^* = i\} \right\} \quad (29)$$

接下来，对式(29)的各部分分别进行推导。首先，推导 P_i^{N1} 的解析表达式， $|f|^2$ 与 $|h_0|^2$ 和 $|g_0|^2$ 相互独立，同时设 $x = |h_0|^2$ ， $y = |g_0|^2$ ， P_i^{N1} 的计算过程为

$$\begin{aligned}
P_i^{N1} = & \Pr \left\{ |f|^2 \geq \frac{\tau_0}{\rho} \right\}. \\
\Pr \left\{ x < \frac{\tau_0}{\rho} - \mu \bar{\alpha} y, y \geq \frac{\tau_i}{\mu \rho (\alpha - \bar{\alpha} \tau_i)} \right\} = & \int_{\frac{\tau_0}{\rho}}^{\infty} \frac{1}{F} e^{-\frac{x}{F}} dx \cdot \\
\int_{a_i}^b \left(\int_0^{\frac{\tau_0}{\rho} - \mu \bar{\alpha} y} \frac{1}{H_0} e^{-\frac{x}{H_0}} dx \right) \frac{1}{G_0} e^{-\frac{y}{G_0}} dy = & \\
e^{-\frac{\tau_0}{\rho F}} \left\{ e^{-\frac{a_i}{G_0}} - e^{-\frac{b}{G_0}} - \frac{H_0 e^{-\frac{\tau_0}{\rho H_0}}}{G_0 \mu \bar{\alpha} - H_0} \left[e^{\left(\frac{\mu \bar{\alpha}}{H_0} - \frac{1}{G_0} \right) b} - e^{\left(\frac{\mu \bar{\alpha}}{H_0} - \frac{1}{G_0} \right) a_i} \right] \right\} & \quad (30)
\end{aligned}$$

其中， $a_i = \frac{\tau_i}{\mu \rho (\alpha - \bar{\alpha} \tau_i)}$ ， $b = \frac{\tau_0}{\mu \rho \bar{\alpha}}$ 。

其次，推导 P_i^{N2} 的表达式，根据信道增益独立性可知 P_i^{N2} 的具体计算步骤为

$$\begin{aligned}
P_i^{N2} = & \Pr \left\{ \frac{P_B |f|^2}{\sigma^2} \geq \tau_0, \frac{P_B |h_0|^2}{\sigma^2} < \tau_0, \right. \\
& \left. \frac{P_C |g_0|^2 \alpha}{P_C |g_0|^2 (1-\alpha) + \sigma^2} < \tau_i \right\} = \\
& \Pr \left\{ |f|^2 \geq \frac{\tau_0}{\rho} \right\} \Pr \left\{ |h_0|^2 < \frac{\tau_0}{\rho} \right\} \cdot \\
& \Pr \left\{ |g_0|^2 < \frac{\tau_i}{\mu \rho (\alpha - \bar{\alpha} \tau_i)} \right\} = \\
& \int_{\frac{\tau_0}{\rho}}^{\infty} \frac{1}{F} e^{-\frac{x}{F}} dx \int_0^{\frac{\tau_0}{\rho}} \frac{1}{H_0} e^{-\frac{x}{H_0}} dx \int_0^{a_i} \frac{1}{G_0} e^{-\frac{x}{G_0}} dx = \\
& e^{-\frac{\tau_0}{\rho F}} \left(1 - e^{-\frac{\tau_0}{\rho H_0}} \right) \left(1 - e^{-\frac{\tau_0}{G_0}} \right) \quad (31)
\end{aligned}$$

最后，计算 $\Pr\{k^* = i\}$ ，根据 X_k 的表达式可知

$$X_k = \begin{cases} \frac{|g_k|^2 \alpha}{\tau_k}, & |h_k|^2 \geq \frac{\tau_0}{\rho} \\ |g_k|^2 \left(\frac{\alpha}{\tau_k} - \bar{\alpha} \right), & |h_k|^2 < \frac{\tau_0}{\rho} \end{cases} \quad (32)$$

根据式(32)， $\Pr\{k^* = i\}$ 可表示为

$$\begin{aligned}
\Pr\{k^* = i\} = & \Pr \left\{ \bigcap_{j \in \{\mathcal{K} \setminus \{i\}\}} X_j < X_i \right\} = \\
& Q_1 \Pr \left\{ |h_i|^2 \geq \frac{\tau_0}{\rho} \right\} + Q_2 \Pr \left\{ |h_i|^2 < \frac{\tau_0}{\rho} \right\} = \\
& Q_1 e^{-\frac{\tau_0}{\rho H_i}} + Q_2 \left(1 - e^{-\frac{\tau_0}{\rho H_i}} \right) \quad (33)
\end{aligned}$$

其中， $Q_1 = \Pr \left\{ \bigcap_{j \in \{\mathcal{K} \setminus \{i\}\}} X_j < \frac{|g_i|^2 \alpha}{\tau_i} \right\}$ ， $Q_2 = \Pr \left\{ \bigcap_{j \in \{\mathcal{K} \setminus \{i\}\}} X_j < |g_i|^2 \left(\frac{\alpha}{\tau_i} - \bar{\alpha} \right) \right\}$ 。首先，计算 Q_1 。对 Q_1 的运算可以等价成一个二重积分，设 $x = |g_i|^2$ ， $y = |g_j|^2$ ，将 x 视为常数先处理有关于 y 的式子，基于此可得

$$\begin{aligned}
 Q_1 &= \int_0^\infty \prod_{j \in \{\mathcal{K} \setminus \{i\}\}} \Pr \left\{ X_j < \frac{x\alpha}{\tau_j} \right\} f_{|g_i|^2}(x) dx = \\
 &\int_0^\infty \prod_{j \in \{\mathcal{K} \setminus \{i\}\}} \left\{ \Pr \left\{ \frac{y\alpha}{\tau_j} < \frac{x\alpha}{\tau_j} \right\} \Pr \left\{ |h_j|^2 \geq \frac{\tau_0}{\rho} \right\} + \right. \\
 &\Pr \left\{ y \left(\frac{\alpha}{\tau_i} - \bar{\alpha} \right) \leq \frac{x\alpha}{\tau_i} \right\} \Pr \left\{ |h_j|^2 < \frac{\tau_0}{\rho} \right\} \left. \right\} \frac{1}{G_i} e^{-\frac{x}{G_i}} dx = \\
 &\int_0^\infty \prod_{j \in \{\mathcal{K} \setminus \{i\}\}} \left[\left(1 - e^{-\frac{\tau_j x}{G_j \tau_j}} \right) e^{-\frac{\tau_0}{\rho H_j}} + \right. \\
 &\left. \left(1 - e^{-\frac{\alpha x}{\tau_j G_j (\alpha/\tau_j - \bar{\alpha})}} \right) \left(1 - e^{-\frac{\tau_0}{\rho H_j}} \right) \right] \frac{1}{G_i} e^{-\frac{x}{G_i}} dx = \\
 &\int_0^\infty \sum_{u=0}^{K-1} \sum_{\substack{S_u \subseteq \{\mathcal{K} \setminus \{i\}\} \\ |S_u|=u}} \left\{ e^{-\sum_{s \in S_u} \frac{\tau_0}{\rho H_s}} \cdot \right. \\
 &\sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \left[(-1)^w e^{-\sum_{t \in S_w} \frac{\tau_0}{\rho H_t}} \right] \cdot \\
 &\left. \sum_{\substack{v=0 \\ |S_v|=v}}^u \sum_{S_v \subseteq S_u} \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \frac{(-1)^{v+w}}{G_i} \Theta(x) \right\} dx \quad (34)
 \end{aligned}$$

其中, $\Theta(x) \triangleq e^{-x \left(\sum_{s \in S_v} \frac{\tau_s}{G_s \tau_i} + \sum_{t \in S_w} \frac{\alpha}{\tau_i G_t} \left(\frac{\alpha - \bar{\alpha}}{\tau_i} \right) + \frac{1}{G_i} \right)}$ 。注意到式(34)的积分运算中只有 $\Theta(x)$ 与 x 有关, 因此针对 $\Theta(x)$ 的积分运算为

$$\begin{aligned}
 \int_0^\infty \Theta(x) dx &= \int_0^\infty e^{-x \left(\sum_{s \in S_v} \frac{\tau_s}{G_s \tau_i} + \sum_{t \in S_w} \frac{\alpha}{\tau_i G_t} \left(\frac{\alpha - \bar{\alpha}}{\tau_i} \right) + \frac{1}{G_i} \right)} dx = \\
 &\frac{1}{\frac{1}{\tau_i} \sum_{s \in S_v} \frac{\tau_s}{G_s \tau_i} + \sum_{t \in S_w} \frac{\alpha}{\tau_i G_t} \left(\frac{\alpha - \bar{\alpha}}{\tau_i} \right) + \frac{1}{G_i}} \quad (35)
 \end{aligned}$$

将式(35)的结果代入式(34), 可得 Q_1 的闭式解为

$$\begin{aligned}
 Q_1 &= \sum_{u=0}^{K-1} \sum_{\substack{S_u \subseteq \{\mathcal{K} \setminus \{i\}\} \\ |S_u|=u}} \left\{ e^{-\sum_{s \in S_u} \frac{\tau_0}{\rho H_s}} \cdot \right. \\
 &\sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \left[(-1)^w e^{-\sum_{t \in S_w} \frac{\tau_0}{\rho H_t}} \right] \cdot \\
 &\left. \sum_{\substack{v=0 \\ |S_v|=v}}^u \sum_{S_v \subseteq S_u} \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \Omega_1 \right\} \quad (36)
 \end{aligned}$$

$$\text{其中, } \Omega_1 \triangleq \frac{(-1)^{v+w}}{G_i \sum_{s \in S_v} \frac{\tau_s}{G_s \tau_i} + G_i \sum_{t \in S_w} \frac{\alpha}{\tau_i G_t} \left(\frac{\alpha}{\tau_i - \bar{\alpha}} \right) + 1}$$

同理可得 Q_2 的闭式解表达式为

$$\begin{aligned}
 Q_2 &= \sum_{u=0}^{K-1} \sum_{\substack{S_u \subseteq \{\mathcal{K} \setminus \{i\}\} \\ |S_u|=u}} \left\{ e^{-\sum_{s \in S_u} \frac{\tau_0}{\rho H_s}} \cdot \right. \\
 &\sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \left[(-1)^w e^{-\sum_{t \in S_w} \frac{\tau_0}{\rho H_t}} \right] \cdot \\
 &\left. \sum_{\substack{v=0 \\ |S_v|=v}}^u \sum_{S_v \subseteq S_u} \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\} \\ |S_w|=w}} \Omega_2 \right\} \quad (37)
 \end{aligned}$$

$$\text{其中, } \Omega_2 \triangleq \frac{(-1)^{v+w}}{G_i \sum_{s \in S_v} \frac{\tau_s \Xi}{G_s \alpha} + G_i \sum_{t \in S_w} \frac{\Xi}{G_t \Xi} + 1}, \quad \Xi = \frac{\alpha}{\tau_i} - \bar{\alpha}$$

在 Q_1 与 Q_2 推导完成后, 代入式(33)即可获得 $\Pr \{k^* = i\}$ 的解析表达式。将式(30)、式(31)与式(33)代入式(29), 即可得到 NOMA 情况下缓存未命中用户 $P_{\text{out}, u_0}^{\text{NOMA}}$ 的表达式。

综上所述, 将式(26)与式(29)代入式(23), 可以得到缓存未命中用户在本文所提调度策略下的中断概率解析表达式为

$$\begin{aligned}
 P_{\text{out}, u_0} &= \sum_{i=1}^K \left\{ (P_i^{N1} + P_i^{N2}) \cdot \right. \\
 &\left[Q_1 e^{-\frac{\tau_0}{\rho H_i}} + Q_2 \left(1 - e^{-\frac{\tau_0}{\rho H_i}} \right) \right] \left. \right\} + \\
 &\left(1 - e^{-\frac{\tau_0}{\rho F}} \right) \left(1 - e^{-\frac{\tau_0}{\rho H_0}} \right) \quad (38)
 \end{aligned}$$

5.3 分集增益分析

当基站 BS 的发送信噪比满足 $\rho \rightarrow \infty$ 时, 可以根据高阶无穷小表达式 $e^{-\frac{\zeta}{\rho}} \simeq 1 - \left(\frac{\zeta}{\rho} \right)$ 将中断概率表达式近似计算得到分集增益, 其中 ζ 表示任意正常数^[18]。基于此, 本文考虑高信噪比情况下, 通过渐近分析得到分集增益, 旨在衡量可靠性随信噪比的改善程度。

对于调度用户 u_k^* , 将高阶无穷小渐近式代

入式(22)可得

$$P_{\text{out},u_k^*} \stackrel{\rho \rightarrow \infty}{\simeq} \prod_{k=1}^K \frac{\tau_k}{\mu \rho G_k} \stackrel{\rho \rightarrow \infty}{\simeq} \frac{1}{\rho^K} \prod_{k=1}^K \frac{\tau_k}{\mu G_k} \propto \frac{1}{\rho^K} \quad (39)$$

因此，调度用户对应的分集增益为 K 阶。

对于缓存未命中用户 u_0 ，首先将高阶无穷小渐近式分别代入式(36)与式(37)，所得近似结果为

$$Q_1 \simeq \sum_{u=0}^{\rho \rightarrow \infty} \sum_{\substack{S_u \subseteq \{\mathcal{K} \setminus \{i\}\} \\ |S_u|=u}} \left\{ \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\}\} \\ |S_w|=w}} (-1)^w \cdot \right. \\ \left. \sum_{v=0}^u \sum_{\substack{S_v \subseteq S_u \\ |S_v|=v}} \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\}\} \\ |S_w|=w}} \Omega_1 \right\} \quad (40)$$

$$Q_2 \simeq \sum_{u=0}^{\rho \rightarrow \infty} \sum_{\substack{S_u \subseteq \{\mathcal{K} \setminus \{i\}\} \\ |S_u|=u}} \left\{ \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\}\} \\ |S_w|=w}} (-1)^w \cdot \right. \\ \left. \sum_{v=0}^u \sum_{\substack{S_v \subseteq S_u \\ |S_v|=v}} \sum_{w=0}^{K-1-u} \sum_{\substack{S_w \subseteq \{\mathcal{K} \setminus \{S_u \cup \{i\}\}\} \\ |S_w|=w}} \Omega_2 \right\} \quad (41)$$

可以看到，在忽略了高阶无穷小项后， Q_1 与 Q_2 近似为常数项，将式(40)与式(41)得到的近似结果以及高阶无穷小渐近式代入式(38)可得

$$P_{\text{out},u_0} \stackrel{\rho \rightarrow \infty}{\simeq} \sum_{i=1}^K \left\{ \left(\frac{\tau_0 [\tau_0 (\alpha - \bar{\alpha} \tau_i) - \bar{\alpha} \tau_i]}{\mu \rho^2 \bar{\alpha} G_0 (\alpha - \bar{\alpha} \tau_i)} + \frac{\tau_0 \tau_i}{\mu \rho^2 G_0 H_0 (\alpha - \bar{\alpha} \tau_i)} \right) \cdot \right. \\ \left. \left[Q_1 \left(1 - \frac{\tau_0}{\rho H_i} \right) + Q_2 \frac{\tau_0}{\rho H_i} \right] + \frac{\tau_0}{\rho F} \frac{\tau_0}{\rho H_0} \right\} \stackrel{\rho \rightarrow \infty}{\simeq} \\ \frac{1}{\rho^2} \left\{ \sum_{i=1}^K \left[\left(\frac{\tau_0 [\tau_0 (\alpha - \bar{\alpha} \tau_i) - \bar{\alpha} \tau_i]}{\mu \bar{\alpha} G_0 (\alpha - \bar{\alpha} \tau_i)} + \frac{\tau_0 \tau_i}{\mu G_0 H_0 (\alpha - \bar{\alpha} \tau_i)} \right) \right] + \frac{\tau_0^2}{F H_0} \right\} \infty \frac{1}{\rho^2} \quad (42)$$

因此，缓存未命中用户的分集增益为 2 阶。

6 仿真分析

本节将使用 MATLAB 对本文中提出的针对无线缓存网络内容交付阶段的用户调度方案的系统

中断性能进行仿真验证。在本文的仿真场景中，缓存服务器 CS 位于坐标 $(0,0)$ ，CS 的服务覆盖半径为 $R_c = 50 \text{ m}$ 。基站 BS 位于坐标 $(-d,0)$ 处， d 表示 BS 与 CS 的相对距离。本节设置的仿真场景下共有 5 个用户，其中，一个用户是缓存未命中用户，该用户无法从 CS 直接请求服务，该用户位于 $(30,10)$ ，请求内容的目标速率为 $r_0 = 1$ ；剩余 $K = 4$ 个用户为缓存命中用户，坐标分别为 $(-35,20)$ 、 $(10,15)$ 、 $(15,-25)$ 以及 $(30,-10)$ ，缓存命中用户请求内容的目标速率分别为 $r_1 = 1.5$ ， $r_2 = 0.8$ ， $r_3 = 2$ ， $r_4 = 1$ 。仿真拓扑如图 3 所示。

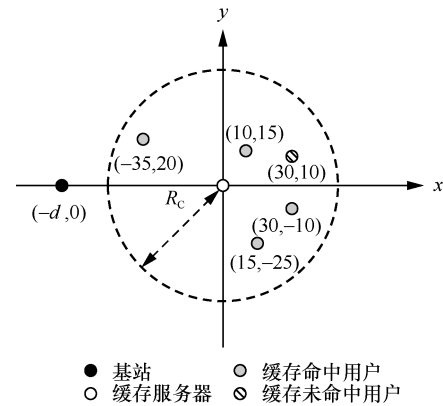


图 3 仿真拓扑

模型中的所有信道增益服从距离相关的指数分布，并设定路径损耗系数 $a = 2$ ，调度用户对应的功率分配因子 $\alpha = 0.8$ 。另外，CS 与 BS 的功率比 $\mu = \frac{P_C}{P_B} = 0.3$ 。调度用户和缓存未命中用户的中断概率理论值分别根据式(22)和式(38)得到，所有的仿真值均从 1×10^7 次蒙特卡罗仿真实验中获得。此外，本文所采用的对比方案如下。

方案 1^[16]。传输策略仍采用本文提出的两阶段自适应 NOMA/OMA 传输方案，用户调度采取基于瞬时 CSI 的调度策略。

方案 2^[15]。传输策略采用传统的 OMA 传输的思想，用户调度采用本文提出的策略。该方案的内容交付的过程分为 3 个阶段：1) 基站 BS 广播内容 x_0 ，CS 与缓存未命中用户 u_0 尝试接收；2) CS 向调度用户使用 OMA 发送内容 x_k ；3) 若在阶段 1) 中 CS 成功解码出了内容 x_0 ，则向缓存未命中用户 u_0 使用 OMA 发送内容 x_0 。而本文中提出的传输方案则分为 2 个阶段，因此，为了保证进行公平的比

较, 该方案中每个内容的目标速率应为两阶段时设置值的 1.5 倍。

方案 3。传输策略采用本文提出的两阶段自适应 NOMA/OMA 传输方案, 用户调度则采取随机选取用户的调度策略。

方案 4^[15]。传输策略为基于 NOMA 的“推送同时分发”方案, 用户调度则采取随机调度。

由于方案 1~方案 4 均需要执行用户调度策略, 因此在实现过程中, 各缓存命中用户 $u_k (k \in \mathcal{K})$ 需要分别根据瞬时 CSI 或生成的随机数来构建虚拟定时器; 此外, 方案 1 和方案 3 中, CS 需要计算是否成功解码 x_0 。因此, 上述 4 个对比方案的计算复杂度均为 $O(K)$ 。

不同方案下调度用户 u_k 与缓存未命中用户 u_0 的中断概率随基站发送信噪比的变化如图 4 和图 5 所示, 其中基站 BS 的坐标为 $(-70,10)$ 。从图 4 和图 5 可以归纳出如下结论。1) u_k 与 u_0 在本文所提方案下的请求内容中断性能的理论值与仿真值完全吻合, 验证了理论分析的正确性。2) 在基站发送信噪比较高的情况下, u_k 与 u_0 的中断概率的理论值与仿真值分别以 ρ^{-4} 与 ρ^{-2} 的速率下降, 证明了对分集增益推导的正确性, 同时也说明了在所提用户调度策略下, 用户 u_k 可以实现满阶的分集增益。此外, 用户 u_0 为 2 阶分集增益的原因是在高发送信噪比下 u_0 采用 MRC 在第一阶段和第二阶段分别接收到内容。3) 4 个对比方案的中断概率在整个发送信噪比范围内均高于所提方案, 证明了在计算复杂度一样的情况下, 所提的高可靠性用户调度策略的有效性。

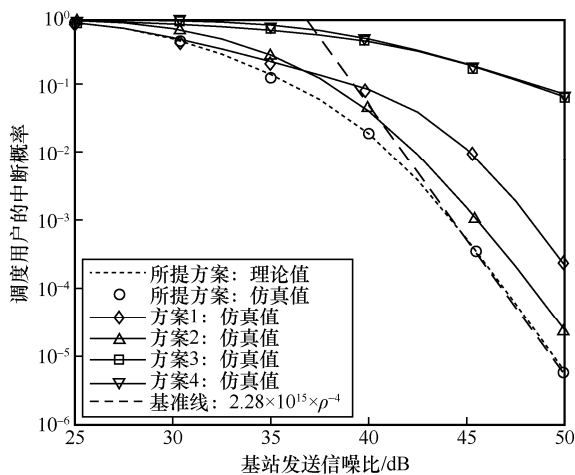


图 4 不同方案下调度用户的中断概率随基站发送信噪比的变化

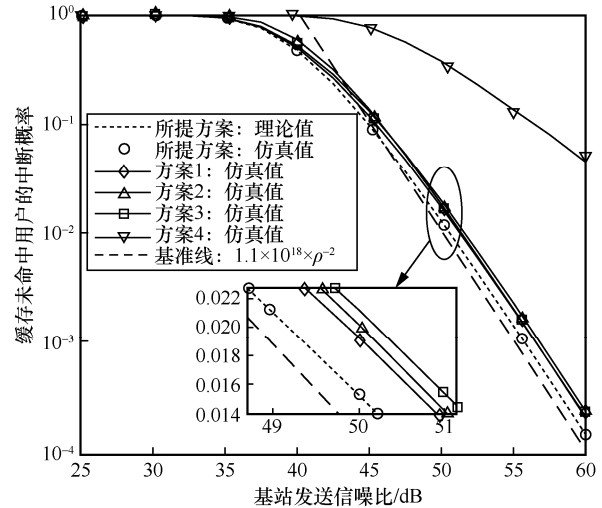


图 5 不同方案下缓存未命中用户的中断概率随基站发送信噪比的变化

不同方案下调度用户 u_k 与缓存未命中用户 u_0 的中断概率随 BS 与 CS 的相对距离 d 的变化分别如图 6 和图 7 所示, 其中基站发送信噪比 $\rho = 45 \text{ dB}$ 。

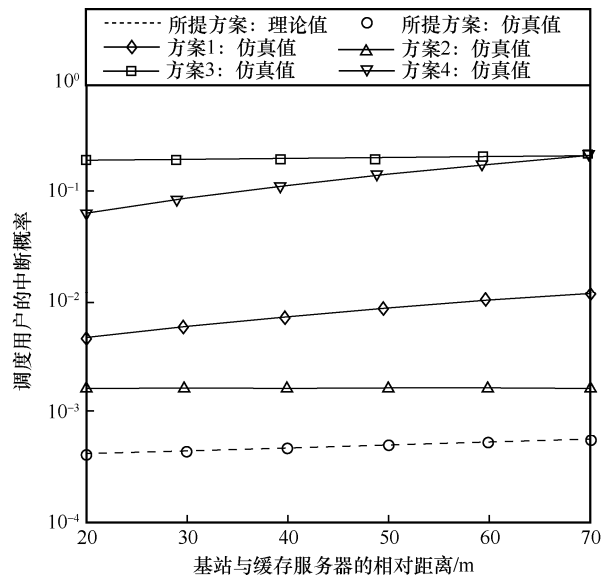


图 6 不同方案下调度用户的中断概率随相对距离的变化

如图 6 所示, 所提方案及基于自适应 NOMA/OMA 的方案 1、方案 3、方案 4 下调度用户的中断概率随着基站与缓存服务器的相对距离 d 单调递增, 而 OMA 方案下调度用户中断概率保持不变, 具体原因如下。在本文所提传输方案中, 调度用户 u_k 将对第一阶段基站广播的内容进行预缓存, 在第二阶段使用 CIC 进行解码。随着 d 的不断增大, u_k 在第一阶段的缓存效果越来越差, 因此会影响第二阶段 CIC 效果, 导致中断概率变大。而

OMA 方案下用户 u_k 仅通过 CS 获取服务，因此与 BS 的位置无关。

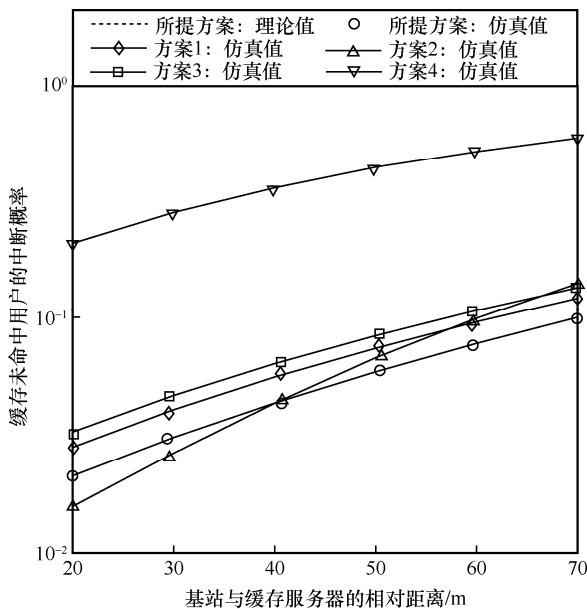


图 7 不同方案下缓存未命中用户的中断概率随相对距离的变化

从图 7 中可以看出，所有方案下缓存未命中用户 u_0 的中断概率随基站与缓存服务器的相对距离的增加而逐渐增大，具体原因如下。虽然用户 u_0 采用 MRC 的形式对内容 x_0 进行解码，但是为了优先保证调度用户 u_k 的传输性，大部分功率分配给了 x_k ，只有少部分功率分配给了 x_0 ，因此第一阶段的广播接收情况对 u_0 的中断概率产生了更大的影响，随着 d 的逐渐增大， u_0 与基站 BS 之间的距离越来越远，因此中断概率也越来越大。另外，当 $d < 40$ m 时，OMA 方案的中断性能优于本文所提方案，而当 $d > 40$ m 时，OMA 方案的性能急剧恶化。这是由于 BS 与 CS、用户 u_0 都保持着很近的距离，且在 OMA 方案下，若 CS 成功缓存了 x_0 ，将以全功率将 x_0 发送给用户 u_0 ，因此中断性能优于本文所提方案。随着 d 的增大，OMA 方案难以满足对应的目标速率要求，因此中断概率不断增大，并且中断性能弱于方案 1、方案 3 和所提方案。

7 结束语

本文面向无线缓存网络的内容分发场景，提出了两阶段自适应 NOMA/OMA 传输方案，针对该方案，设计了一种高可靠性用户调度策略。通过所提策略从其他缓存命中用户中选出调度用户，与缓存

未命中用户组成 NOMA 用户对，由缓存服务器对该用户对进行联合分发，在保证调度用户的可靠性的前提下尽可能地降低缓存未命中用户发生中断的可能性。为了验证所提方案的系统中断性能，分别推导了调度用户与缓存未命中用户的中断概率闭式解表达式，并且在高信噪比的条件下推导了与中断概率对应的分集增益表达式。理论结果表明，调度用户可获得满空间分集增益，缓存未命中用户通过 MRC 也能获得 2 阶分集增益。仿真结果表明，本文所提传输方案与调度策略具有较好的中断性能。

参考文献：

- [1] WANG X F, CHEN M, TALEB T, et al. Cache in the air: exploiting content caching and delivery techniques for 5G systems[J]. IEEE Communications Magazine, 2014, 52(2): 131-139.
- [2] GOLREZAEI N, MOLISCH A F, DIMAKIS A G, et al. Femtocaching and device-to-device collaboration: a new architecture for wireless video distribution[J]. IEEE Communications Magazine, 2013, 51(4): 142-149.
- [3] BASTUG E, BENNIS M, DEBBAH M. Living on the edge: the role of proactive caching in 5G wireless networks[J]. IEEE Communications Magazine, 2014, 52(8): 82-89.
- [4] ZHAO N, LIU X N, YU F R, et al. Communications, caching, and computing oriented small cell networks with interference alignment[J]. IEEE Communications Magazine, 2016, 54(9): 29-35.
- [5] DAI L L, WANG B C, DING Z G, et al. A survey of non-orthogonal multiple access for 5G[J]. IEEE Communications Surveys & Tutorials, 2018, 20(3): 2294-2323.
- [6] DAI L L, WANG B C, YUAN Y F, et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends[J]. IEEE Communications Magazine, 2015, 53(9): 74-81.
- [7] DING Z G, LIU Y W, CHOI J H, et al. Application of non-orthogonal multiple access in LTE and 5G networks[J]. IEEE Communications Magazine, 2017, 55(2): 185-191.
- [8] XIANG L, NG D W K, GE X H, et al. Cache-aided non-orthogonal multiple access: the two-user case[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(3): 436-451.
- [9] GURUGOPINATH S, SOFOTASIOS P C, AL-HAMMADI Y, et al. Cache-aided non-orthogonal multiple access for 5G-enabled vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(9): 8359-8371.
- [10] SHEN K Z, ALHARBI T E A, SO D K C. Cache-aided device-to-device non-orthogonal multiple access[C]//Proceedings of 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). Piscataway: IEEE Press, 2020: 1-6.
- [11] SHEN K Z, SO D K C. Power allocation for D2D NOMA in cache-aided networks[C]//Proceedings of IEEE 94th Vehicular Technology Conference (VTC2021-Fall). Piscataway: IEEE Press, 2021: 1-6.
- [12] FU Y R, WEN W L, ZHAO Z Y, et al. Dynamic power control for NOMA transmissions in wireless caching networks[J]. IEEE Wireless Communications Letters, 2019, 8(5): 1485-1488.
- [13] ZHAO Z Y, XU M F, XIE W L, et al. Coverage performance of NOMA in wireless caching networks[J]. IEEE Communications Letters, 2018, 22(7): 1458-1461.

[14] YIN Y, LIU M, GUI G, et al. QoS-oriented dynamic power allocation in NOMA-based wireless caching networks[J]. IEEE Wireless Communications Letters, 2021, 10(1): 82-86.

[15] DING Z G, FAN P Z, KARAGIANNIDIS G K, et al. NOMA assisted wireless caching: strategies and performance analysis[J]. IEEE Transactions on Communications, 2018, 66(10): 4854-4876.

[16] LV L, YE Q, DING Z G, et al. Multi-antenna two-way relay based cooperative NOMA[J]. IEEE Transactions on Wireless Communications, 2020, 19(10): 6486-6503.

[17] YANG L, CHEN J, ZHANG H L, et al. Cooperative wireless multicast: performance analysis and time allocation[J]. IEEE Transactions on Vehicular Technology, 2016, 65(7): 5810-5819.

[18] COLDSMITH A. Wireless communications[M]. Cambridge: Cambridge University Press, 2005.

[19] YUE X W, LIU Y W, KANG S L, et al. Exploiting full/half-duplex user relaying in NOMA systems[J]. IEEE Transactions on Communications, 2018, 66(2): 560-575.



任萌萌（1996-），女，陕西咸阳人，西安电子科技大学博士生，主要研究方向为非正交多址接入、移动边缘计算等。



贺冰涛（1991-），男，陕西西安人，博士，西安电子科技大学讲师，主要研究方向为协作通信、物理层安全和非正交多址接入。

[作者简介]



杨龙（1988-），男，陕西潼关人，博士，西安电子科技大学副教授、博士生导师，主要研究方向为非正交多址接入、移动边缘计算、协作通信等。



周雨晨（1991-），女，陕西咸阳人，博士，西安电子科技大学副教授、硕士生导师，主要研究方向为物联网、信息物理系统、无线网络虚拟化等。



李函洋（1997-），男，河北唐山人，西安电子科技大学硕士生，主要研究方向为非正交多址接入、无线缓存等。



陈健（1968-），男，江苏如东人，博士，西安电子科技大学教授、博士生导师，主要研究方向为物理层安全、无线传感网络、非正交多址接入等。