



文章编号:1001-9081(2019)09-2664-05

DOI:10.11772/j.issn.1001-9081.2019030509

移动边缘计算中一种贪心策略的内容卸载方案

袁培燕*, 蔡云云

(河南师范大学 计算机与信息工程学院, 河南 新乡 453007)

(*通信作者电子邮箱 peiyan@htu.cn)

摘要: 基于移动边缘计算的内容卸载技术可以有效降低骨干网络的流量压力, 提升终端用户体验。针对终端用户与小基站之间的异质接触率, 设计了一种贪心策略的内容卸载方案。首先, 将内容最优卸载问题转化为内容最大投递率问题; 其次, 证明最大投递率问题满足子模性, 在此基础上, 采用贪心算法部署内容, 该算法可以以概率($1 - 1/e$) 保证其最优化; 最后, 细致分析了内容流行度指数以及缓存大小对不同卸载方案的影响。实验结果表明, 所提方案提高了内容投递率同时降低了内容传输时延。

关键词: 移动边缘计算; 投递率; 内容卸载; 子模性; 贪心算法

中图分类号: TP393.01 **文献标志码:** A

Content offloading scheme of greedy strategy in mobile edge computing system

YUAN Peiyan*, CAI Yunyun

(College of Computer and Information Engineering, Henan Normal University, Xinxiang Henan 453007, China)

Abstract: Content offloading technology based on mobile edge computing can effectively reduce the traffic pressure on the backbone network and improve the end user's experience. A content offloading scheme of greedy strategy was designed for the heterogeneous contact rate between end users and small base stations. Firstly, the content optimal offloading problem was transformed into the content maximum delivery rate problem. Secondly, the maximum delivery rate problem was proved to satisfy the submodularity. On this basis, the greedy algorithm was used to deploy the content. The algorithm was able to guarantee its optimality with the probability ($1 - 1/e$). Finally, the impacts of content popularity index and cache size on different offloading schemes were analyzed in detail. The experimental results show that the proposed scheme improves the content delivery rate and reduces the content transmission delay at the same time.

Key words: mobile edge computing; delivery rate; content offloading; submodularity; greedy algorithm

0 引言

近年来, 随着无线通信技术的快速发展和便携式设备的大规模普及, 逐渐形成了人、机物互联互通的泛在网络环境。与此同时, 大规模的终端设备进行数据传输, 消耗了大量的带宽、能量和时间, 终端用户的服务质量 (Quality of Service, QoS) 急剧下降。移动边缘计算 (Mobile Edge Computing, MEC) 技术通过将存储、计算和通信集成到宏基站 (Base Station, BS) 或微基站 (Small Base Station, SBS) 等更加贴近用户的地方, 在一定程度上降低了时延、提高了网络运营效率^[1-3], 同时实现资源在终端、边缘和云中心的 3 层级优化利用, 更好满足用户体验质量需求^[4-5]。

内容卸载作为移动边缘计算的一项典型应用, 吸引了学术界和工业界的广泛关注。文献[6]提出利用住宅 WiFi 接入点 (Access Point, AP) 来卸载数据。文献[7]将蜂窝网络中的车辆通信业务卸载到车辆到车辆 (Vehicle-to-Vehicle, V2V) 的行进路径中。基于卸载控制方案移动边缘计算内的软件定义网络 (Software-Defined Network inside the Mobile Edge Computing, SDNi-MEC) 服务架构的分析表明, 当车辆密度处于中间值时, 它在蜂窝网络链路和 V2V 路径中具有更大的吞

吐量。内容卸载之所以能够提升网络性能, 是因为它将缓存的内容从距离用户较远的地方 (云中心) 卸载到距离用户终端较近的地方 (宏基站或微基站), 可以利用短距离通信技术提升系统容量。文献[8]基于单个节点的缓存大小、内容数量和网络节点数量, 研究了网络容量的扩展律。证明即使网络中节点数量增加, 具有缓存的无线自组织网络的容量也可以保持一个常量。

本文主要考虑将网络中的部分内容预先卸载到 SBS 上, 进而更好地服务用户。由于每一个微基站缓存大小及覆盖范围有限, 于是一个性能良好的卸载方案就变得至关重要。与相关工作不同的是, 本文在将网络中的内容卸载到 SBS 的过程中, 利用用户终端与 SBS 之间的异质接触率, 通过贪心算法对内容进行卸载。具体贡献如下:

1) 在移动边缘计算系统中将内容最优卸载问题转化为最大投递率问题, 确保 SBS 所存放的内容可以满足大部分用户的需求;

2) 证明了最大投递率问题满足子模性, 在此基础上, 设计了一种贪婪算法求解最优 SBS 选择问题;

3) 分析了缓存大小和流行度指数对于投递率以及传输时延的影响, 仿真结果表明基于贪心算法的缓存策略与经典

收稿日期:2019-03-27;修回日期:2019-05-14;录用日期:2019-05-15。

基金项目:国家自然科学基金资助项目(U1804164, U1404602);河南省科技攻关基金资助项目(172102210341)。

作者简介:袁培燕(1978—),男,河南邓州人,副教授,博士,CCF 高级会员,主要研究方向:移动网络、无线通信; 蔡云云(1994—),女,河南安阳人,硕士研究生,主要研究方向:移动边缘计算。



的缓存策略相比,投递率与传输时延均有明显改善。

1 相关工作

内容卸载是移动网络的一个重要应用。内容卸载问题与缓存问题紧密相关,增加缓存以及进行内容卸载都可以使网络性能得到提升。即使每一个内容只缓存一个备份,也可以使网络性能得到提升。然而在现实中,不同的文件有不同的流行度,流行度高的文件在网络中有更高的概率被用户终端请求,流行度低的文件可能被较少的用户请求,甚至有些文件在网络中不被请求。在流量使用高峰期,请求流行度高的内容的用户终端可能会出现排队现象,造成网络延迟。流行度较低的内容可能没有被用户终端请求,造成资源浪费。文献[9]按照内容的流行度对内容进行存储。有较高流行度的内容在网络中有较多的备份,较低流行度的内容在网络中有较少的备份。仿真结果表明,按照流行度缓存计算出的容量高于所有内容均匀缓存的容量。这是因为流行度高的内容在网络中的备份数多,用户终端会相对容易找到自己请求的文件,传输距离减小,容量增大。

依照流行度进行缓存固然重要,缓存策略也很重要。当一个BS将内容卸载到SBS时,无法确认将内容缓存在哪些SBS中更容易满足用户需求。最优卸载问题可以被理解为预缓存问题。近年来已有一些关于预缓存问题的研究。Liu等^[10]研究了最佳内容放置,基于概率缓存框架,采用随机几何理论推导出闭形的成功卸载概率,并制定了缓存概率优化问题。文献[11]利用设备到设备的通信技术,将公共内容分发给在下载过程期间协作的一组移动设备。基站将内容的不同组块单播到选定的移动设备,移动设备使用多跳协作,同时保持对移动设备的能量消耗的公平性约束。通过将最优蜂窝卸载问题表述为混合整数线性规划问题,分析其算法复杂性。实验结果表明,即使只有很少一部分移动设备用于合作,也可以显著的增益系统性能。Han等^[12]设计了一种有偏好的学习算法来跟踪用户不断变化的兴趣,然后通过预测WiFi网络连接的持续时间,优化了何时预取以最大化用户体验问题,同时降低预取成本。文献[13]提出基于推理的社交网络内容预取器EarlyBird,使用社交数据特定信号,以便在用户使用之前检索新闻提要以及相关链接和多媒体。提出的基于回归的内容预测模型能够在55%的时间内估计用户感兴趣的内容。文献[14]提出了一种基于动作的预取解决方案Precog,用于时移WiFi卸载。与需要并发蜂窝和WiFi连接的先前卸载解决方案不同,Precog通过时移WiFi卸载内容。文献[15]通过带宽有限的两层异构网络中的带宽分配和缓存放置来最小化平均文件传输延迟。对于联合带宽分配和高速缓存放置问题,首先给出最佳带宽分配策略。然后,通过替换最佳带宽分配,将原始问题转换成关于高速缓存策略的等效问题。最后为了解决等效非凸问题,提出了一种低复杂度的迭代算法来获得次优解。文献[16]将延迟最小化作为优化目标并将其表示为整数规划问题,通过动态规划有效地分配资源。结合以上工作不难发现,无论是卸载问题还是缓存问题,其目的都是将内容和资源合理分配,从而提升网络性能。

2 系统架构

移动边缘计算通常由边缘设备(如智能手机、物联网设

备、智能车等),边缘云和远端云(或大型云计算中心、大云)组成。边缘云是部署在移动基站上的小规模云计算中心,更加专注于局部,将计算任务在接近数据源的地方执行,可以有效减小计算系统的延迟,减小内容传输带宽,缓解云计算中心压力并能够保护数据安全和隐私^[17]。当边缘设备的处理能力无法满足自身需求时,可以通过无线网络将密集型任务和海量数据迁移到边缘云处理;当边缘云不能满足边缘设备的要求时,任务和数据可以迁移到远端云处理。在本文中,边缘设备为了更好地满足用户终端的需求并降低自身压力,将内容卸载在SBS上。然后,为了最大化增益,用贪心算法求解内容最佳卸载问题。

假设一个边缘网络包括一个BS, n 个SBS。所有SBS的集合可以表示为 $V = \{v_1, v_2, \dots, v_n\}$ 。整个边缘网络中有 m 个独立的内容,记作 $F = \{F_1, F_2, \dots, F_m\}$ 。为了方便计算,假设每一个内容的大小是1b。每一个SBS空间大小是 s b,即每一个SBS可以缓存 s 个内容,其中 $s \leq m$ 。为了减少等待时间并且将SBS中的缓存空间充分利用,每个内容在每一个SBS中仅具有一个缓存。另外,假设每个内容都有一个生命周期 T ,剩余时间越长表示该内容被移动终端收到的概率越大。

用户终端到SBS的传输距离小于用户终端到BS之间的传输距离,如果用户终端尽可能多地在SBS请求内容,而不是向BS请求内容,该内容请求将会更快得到响应。由于用户终端是移动的并且每一个SBS有不同的位置和覆盖范围,因此用户终端跟每一个SBS的接触率不同。同样的内容在不同的SBS中有不同的投递率(投递率是用于评估算法性能的重要标准之一,其可以直观地反映网络传输的可靠性。当源节点发送的内容数量不变时,目的节点成功接收的内容越多,内容的投递率就越高)。另外,每一个SBS的缓存空间有限,只能缓存部分内容。这就使得对每一个内容进行合理分配尤为重要。为了使传输更加可靠、传输时延更小,应该将内容卸载到使投递率尽可能大的SBS中。因为不同的内容有不同的流行度,所以将优化问题表示为:

$$\max \left(\sum_{i=1}^m p_i d_i \right) \quad (1)$$

s.t.

$$\begin{cases} A_i \leq n \\ \sum_{i=1}^m A_i \leq ns \end{cases} \quad (2)$$

其中: p_i 指第 i 个内容的流行度, d_i 指第 i 个内容的投递率。第一个限制条件表示每一个内容的数量不能超过SBS的数量;第二个限制条件表示缓存的所有内容的数量不能超过所有SBS的缓存大小。

在解决优化问题之前,需要计算第 i 个内容的流行度。假设内容的流行度遵循Zipf分布,并且第 i 个内容的流行度与 $1/i^\gamma$ 成正比。于是可以将第 i 个内容的流行度表示为:

$$p_i = \frac{1}{i^\gamma H_{m,\gamma}} = 1 / \left(i^\gamma \sum_{j=1}^m \frac{1}{j^\gamma} \right) \quad (3)$$

满足 $0 \leq p_i < 1$, $\sum_{i=1}^m p_i = 1$, 其中, $H_{m,\gamma}$ 是广义谐波数。至此,可以通过解决最大投递率问题来解决内容的最佳卸载问题。



3 问题解决

本章中主要通过最大投递率问题解决内容的最佳卸载问题。首先研究用公式表示最大投递率,其次证明最大投递率问题满足子模性,最后用贪心策略解决内容最佳卸载问题。

3.1 子模性

假设 β_j 表示用户终端与第 j 个SBS之间接触概率,存放在第 j 个SBS中的内容在时间 T 内没有被目的用户终端接收到的概率为 $e^{-\beta_j T}$,则目的用户终端接收到的概率可以表示为 $1 - e^{-\beta_j T}$ 。令集合 A_i 表示当前存储第 i 个内容的SBS的集合。 $A = \|A_i\|$ 表示缓存第 i 个内容的SBS的数量。假设存在一个集合 B_i ,满足 $A_i \subseteq B_i$, $B = \|B_i\|$ 。由于第 i 个内容的投递率 d_i 可以表示为 $1 - e^{-\sum_{j \in A_i} \beta_j T}$,结合式(3),则式(1)可以表示为:

$$\sum_{i=1}^m p_i \left(1 - e^{-\sum_{j \in A_i} \beta_j T} \right) \quad (4)$$

s. t.

$$\begin{cases} A_i \leq n \\ \sum_{i=1}^m A_i \leq ns \end{cases} \quad (5)$$

定理1 最大投递率问题满足子模性。

证明 当有一个新的SBS加入时,即基站想要将第 i 个文件放置在一个SBS,存放第 i 个内容的SBS的集合可以表示为 $A'_i = A_i \cup \{v\}$, $B'_i = B_i \cup \{v\}$ 。

对于集合 A_i ,加入 v 后,投递率的增量为:

$$\begin{aligned} p_i \left(1 - e^{-\sum_{j \in A'_i} \beta_j T} \right) - p_i \left(1 - e^{-\sum_{j \in A_i} \beta_j T} \right) = \\ p_i \left(e^{-\sum_{j \in A'_i} \beta_j T} - e^{-\sum_{j \in A_i} \beta_j T} \right) = p_i e^{-\sum_{j \in A'_i} \beta_j T} (1 - e^{-\beta_i T}) \end{aligned} \quad (6)$$

同理,对于集合 B_i ,投递率的增量为:

$$\begin{aligned} p_i \left(1 - e^{-\sum_{j \in B'_i} \beta_j T} \right) - p_i \left(1 - e^{-\sum_{j \in B_i} \beta_j T} \right) = \\ p_i e^{-\sum_{j \in B'_i} \beta_j T} (1 - e^{-\beta_i T}) \end{aligned} \quad (7)$$

因为 $A_i \subseteq B_i$,可以得出:

$$p_i e^{-\sum_{j \in A'_i} \beta_j T} (1 - e^{-\beta_i T}) \geq p_i e^{-\sum_{j \in B'_i} \beta_j T} (1 - e^{-\beta_i T}) \quad (8)$$

所以,内容的投递率满足子模性,

证毕。

3.2 贪婪算法

算法1 基于贪婪算法的内容卸载流程。

输入:所有的SBS的集合 V ;

输出:能够使第 i 内容的投递率最大的SBS的集合 C 。

- 1) 计算出将第 i 内容放入每一个SBS的投递率并排序
- 2) $D = \{i\}$
- 3) $V_0 = V$
- 4) $V = V - \{i\}$
- 5) while $\sum_{i=1}^m a_i \leq ns, a_i \leq n$ do
- 6) for each node $r \in V \setminus C$ do
- 7) $r \leftarrow \arg \max_{r \in V \setminus C} (f(C \cup \{r\}) - f(C))$
- 8) end for
- 9) $C = C \cup \{r\}$
- 10) $V = V - \{r\}$
- 11) end while
- 12) return C

由于投递率满足子模性,因此存在一个贪心算法以(1 -

$1/e$)的概率保证式(4)的最优性。贪心算法是一种改进了的分级处理的方法,其核心是根据问题选取一种量度标准。然后将多个输入排成这种量度标准所要求的顺序,按这种顺序一次输入一个量。如果这个输入和当前已构成在这种量度标准下的部分最佳解合并在一起不能产生一个可行解,则此输入不能加入到该部分解中。这种能够得到某种量度意义下最优解的分级处理方法称为贪心算法。

本文贪心算法的核心是将投递率作为一种度量标准。按照将第 i 个内容放置到SBS的投递率等级进行排序。基站按照这个顺序将内容依次放入SBS。如果新加入的内容放置的SBS和当前已经放入第 i 内容的SBS的集合一起可以使投递率最高,则第 i 内容放入该SBS,并将该SBS加入存放第 i 个内容的SBS的集合 C 内。当有一个内容需要加入SBS时重复上述过程,直到所有SBS没有空闲缓存空间或所有内容均被缓存。

4 实验仿真

本章分析流行度指数以及缓存的大小对于投递率和传输时延的影响,并将基于贪心算法(Greedy algorithm)进行缓存的投递率和传输时延与经典方案进行比较:随机缓存(Random)^[18]、根据流行度缓存(Popular)^[7]和根据阈值缓存(Pop-uppop)^[19]。随机缓存表示每个节点对每个内容按照一定的概率随机缓存,不考虑内容的流行度。根据流行度缓存的方案是指每一个内容在节点中的备份数由内容的流行度决定,若内容流行度较高,则该内容在节点中有较多的备份数;若内容的流行度较低,则缓存该内容较少。根据阈值缓存方案指流行度最高的前 k 个内容在每一个SBS中都有一个备份,第 $k+1$ 个内容至最后一个内容在所有SBS组成的网络中最多仅有一个备份。

图1(a)与图1(b)中分析了流行度指数以及缓存大小对投递率的影响。在图1(a)中,缓存大小设置为10。从图中可以看出随着流行度指数的不断增大,根据贪心算法进行缓存的投递率在不断增大。在流行度指数接近于2.0时,投递率接近1。基于阈值的缓存方案和基于流行度进行缓存方案的投递率也会随着流行度指数的增大而增大,但是投递率的值在流行度指数增长期间,始终没有超过基于贪心算法进行缓存的投递率的值。随机缓存有很多不确定性,其投递率也存在很多不确定性。为了能够较好地反映基于随机缓存的投递率随着流行度指数的变化情况,进行了4次仿真实验,并取均值。从图1(a)中可以看出,随机缓存的投递率平均值很小。在流行度指数变化期间,投递率一直处于低于0.1的状态。在图1(b)中,流行度指数设置为1.0。几种方案随着缓存大小的变大,变化不是太明显。但是基于贪心算法进行缓存的方案,投递率仍然是单调上升的趋势,并且其值大于其他方案下的投递率的值。结合图1中的两个图可以发现,基于贪心算法进行缓存的方案跟其他方案相比显示出一定的优越性。

图2(a)和与图2(b)中分析了流行度指数以及缓存区的大小对传输时延的影响。在图2(a)中,将缓存大小设置为10,传输时延的仿真图与投递率的变化趋势几乎是对称的。在图2(a)中,基于贪心算法进行缓存的传输时延与其他方案的传输时延随着流行度指数的增大而快速下降,但是基于贪心算法进行缓存的传输时延的下降趋势大于基于阈值的缓存



和基于流行度缓存的下降趋势。随机缓存的传输时延并没有随着流行度指数的变化而下降,在整个流行度指数的变化期间一直保持接近10的状态。图2(b)中,流行度指数设置为1.0,并且随着缓存变大,传输时延的变化趋势都不是太大。基于贪心算法缓存的传输时延最小。从传输时延的变化趋势可以看出,传输时延在减小到一定的值的时候,变化不会太明显。结合实际,用户与SBS进内容行输,总会有一定的传输时延。这是因为SBS与用户终端之间会有一定的传输距离,且内容处理都需要时间。结合投递率以及传输时延,跟其他经典的方案相比,基于贪心算法进行缓存的方案提升了投递率,减小了传输时延。

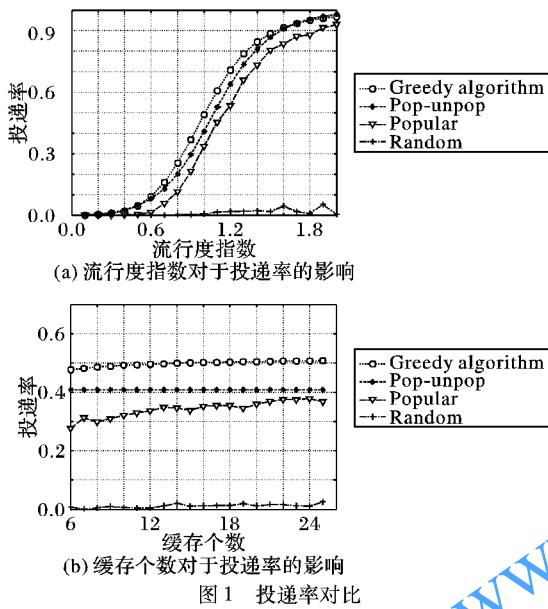


Fig. 1 Comparison of delivery rate

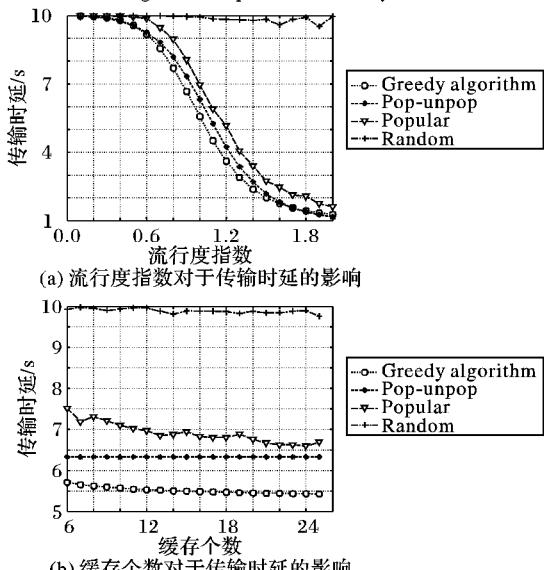
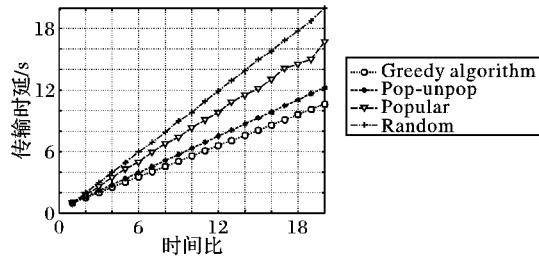


Fig. 2 Comparison of transmission delay

图3是4种缓存方案下的投递率随着时间比(BS 到用户终端与 SBS 到用户终端传输的时间比)增大的变化情况。将 k 、流行度指数以及缓存大小分别设置为 30、10 和 1.0。时间比变大表示用户请求的内容由 BS 提供服务的比例大于由 SBS 提供服务的比例。因为 BS 到用户终端的距离大于 SBS

到用户终端的距离,所以用户的请求由 BS 提供服务的传输时延比 SBS 提供服务的传输时延大。时间比对于传输时延的影响再一次证明在靠近用户终端进行缓存的重要性,也证明了一个性能良好的卸载方案的重要性。

图3 时间比对于传输时延的影响
Fig. 3 Effect of time ratio on transmission delay

5 结语

本文针对移动边缘计算中内容的最佳卸载问题,提出用贪婪算法解决。将内容卸载问题转化为最大投递率问题,并证明最大投递率问题满足子模性,因此用贪心算法可以概率 $(1 - 1/e)$ 保证其最优性。仿真结果表明,与一些经典方案相比,提出的方案使内容投递率与传输时延性能均得到了较大改善。

参考文献 (References)

- [1] BARBAPOSSA S, CECI E, MERLUZZI M, et al. Enabling effective mobile edge computing using millimeterwave links [C]// Proceedings of 2017 IEEE International Conference on Communications Workshops. Piscataway, NJ: IEEE, 2017: 367 – 372.
- [2] BELLI D, CHESSA S, FOSCHINI L, et al. A social-based approach to mobile edge computing [C]// Proceedings of the 2018 IEEE Symposium on Computers and Communications. Piscataway, NJ: IEEE, 2018: 292 – 297.
- [3] ROMAN R, LOPEZ J, MAMBO M. Mobile edge computing, Fog et al.: a survey and analysis of security threats and challenges [J]. Future Generation Computer Systems, 2018, 78: 680 – 698.
- [4] 邓晓衡, 关培源, 万志文, 等. 基于综合信任的边缘计算资源协同研究[J]. 计算机研究与发展, 2018, 55(3): 449 – 477.
(DENG X H, GUAN P Y, WAN Z W, et al. Integrated trust based resource cooperation in edge computing [J]. Journal of Computer Research and Development, 2018, 55(3):449 – 477.)
- [5] GUAN P, DENG X, LIU Y, et al. Analysis of multiple clients' behaviors in edge computing environment [J]. IEEE Transactions on Vehicular Technology, 2018, 67(9): 9052 – 9055.
- [6] POULARAKIS K, IOSIFIDIS G, PEFKIANAKIS I, et al. Mobile data offloading through caching in residential 802.11 wireless networks [J]. IEEE Transactions on Network and Service Management, 2016, 13(1): 71 – 84.
- [7] HUANG C, CHIANG M, DAO D, et al. V2V data offloading for cellular network based on the Software Defined Network (SDN) inside Mobile Edge Computing (MEC) architecture [J]. IEEE Access, 2018, 6: 17741 – 17755.
- [8] QIU L, CAO G. Cache increases the capacity of wireless networks [C]// Proceedings of the 35th Annual IEEE International Conference on Computer Communications. Piscataway, NJ: IEEE, 2016: 1 – 9.
- [9] QIU L, CAO G. Popularity-aware caching increases the capacity of



- wireless networks [C]// Proceedings of the 2017 IEEE Conference on Computer Communications. Washington, DC: IEEE Computer Society, 2019: 1–9.
- [10] LIU D, YANG C. Optimal content placement for offloading in cache-enabled heterogeneous wireless networks [C]// Proceedings of the 2016 IEEE Global Communications Conference. Piscataway, NJ: IEEE, 2016: 1–6.
- [11] AL-KANJ L, POOR H V, DAWY Z. Optimal cellular offloading via device-to-device communication networks with fairness constraints [J]. *IEEE Transactions on Wireless Communications*, 2014, 13(8): 4628–4643.
- [12] HAN J, LI X, JUN T, et al. Network agile preference-based prefetching for mobile devices [C]// Proceedings of the 2014 IEEE 33rd International Performance Computing and Communications Conference. Piscataway, NJ: IEEE, 2014: 1–8.
- [13] WANG Y, LIU X, CHU D, et al. EarlyBird: mobile prefetching of social network feeds via content preference mining and usage pattern analysis [C]// Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM, 2015: 67–76.
- [14] SANADHYA S, MORAVAPALLE U P, KIM K-H, et al. Precog: action-based time-shifted prefetching for Web applications on mobile devices [EB/OL]. [2018-12-20]. http://gnan.ece.gatech.edu/archive/precog_hotweb_17.pdf.
- [15] YANG Z H, PAN C H, PAN Y J, et al. Cache placement in two-tier HetNets with limited storage capacity: cache or buffer? [J]. *IEEE Transactions on Communications*, 2018, 66(11): 5415–5429.
- [16] LUO J, DENG X, ZHANG H, et al. Ultra-low latency service provision in edge computing [C]// Proceedings of the 2018 IEEE International Conference on Communications. Piscataway, NJ: IEEE, 2018: 1–6.
- [17] 施巍松, 张星洲, 王一帆, 等. 边缘计算: 现状与展望 [J]. *计算机研究与发展*, 2019, 56(1): 69–89. (SHI W S, ZHANG X Z, WANG Y F, et al. Edge computing: state-of-the-art and future directions [J]. *Journal of Computer Research and Development*, 2019, 56(1): 69–89.)
- [18] WEN W, CUI Y, ZHENG F C, et al. Random caching based cooperative transmission in heterogeneous wireless networks [J]. *IEEE Transactions on Communications*, 2018, 66(7): 2809–2825.
- [19] AO W C, PSOUNIS K. Distributed caching and small cell cooperation for fast content delivery [C]// Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM, 2015: 127–136.

This work is partially supported by the National Natural Science Foundation of China (U1804164, U1404602), the Science and Technology Foundation of Henan Province (172102210341).

YUAN Peiyun, born in 1978, Ph. D., associate professor. His research interests include mobile network, wireless communication.

CAI Yunyun, born in 1994, M. S. candidate. Her research interests include mobile edge computing.