



移动边缘计算中的内容分发加速策略

柳 兴^{1,2*}, 杨 震¹, 王新军², 朱 恒²

(1.北京邮电大学 计算机学院, 北京 100876;

2. 中国联合网络通信有限公司 联通(湖南)产业互联网研究院, 湖南 长沙 410014)

(*通信作者电子邮箱 liuxing_bupt@qq.com)

摘 要: 针对移动边缘计算中的内容分发加速问题, 考虑 MEC 服务器存储空间受限对内容缓存的影响, 以移动用户获取对象延时为优化目标, 根据用户群对不同对象的兴趣的差异, 提出了一种基于兴趣的内容分发加速策略。该策略根据 MEC 服务器的存储空间、移动用户群对不同对象的兴趣以及对象的文件大小, 选择性地在 MEC 服务器上缓存对象, 并及时对 MEC 服务器上缓存的对象进行更新, 最大限度的满足移动用户群的内容需求。仿真结果表明, 本文所提策略具有良好的收敛性能, 所提策略的缓存命中率相对稳定且明显优于现有策略; 当系统运行达到稳定后, 较现有策略, 用户获取对象数据的时延可减少 20%。

关键词: 5G; 移动边缘计算; 内容分发; 缓存; 加速; 同步

中图分类号: TP302

文献标志码: A

Content distribution speedup strategy in mobile edge computing

Liu Xing^{1,2*}, Yang Zhen¹, Wang Xingjun², Zhu Heng²

(1. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Uincom (Hunan) Industry Internet Research Institute, ChinaUnicom, Hunan Changsha 41004, China)

Abstract: The MEC server can reduce the delay for mobile users to get files by caching, but its storage space limits performance gains. Focusing on this problem, an Interest-based Content Distribution Acceleration Strategy (ICDAS) is proposed in this paper. The proposed strategy selectively cache objects on MEC servers and timely update the object content to maximize the content requirements of mobile user groups. The selection process considers the storage space of the base station, the interest of the mobile user group on different objects, and the file size of the objects. The experimental results show that the proposed strategy has good convergence performance. The cache hit ratio of the proposed strategy is relatively stable and significantly better than the existing strategy. When the system runs stably, compared with the existing strategy, the delay for users to obtain the object data can be reduced by 20%.

Keywords: 5G; EMC; content distribution; cache; speedup; synchronization

0 引言

随着 5G 的商用, 流量数据应用(如: 人人视频、爱奇艺、快手、抖音等)必然愈发普及^[1]。边缘计算将以更快的网络服务响应能力, 受到各类移动用户的青睐^[2]。有研究表明, 未来流量数据将是视频流量为主^[3]。因此, 如何合理

的设计移动边缘计算(Mobile Edge Computing, MEC)的缓存策略是值得研究的课题。

目前, 已经有多个文献对 MEC 中的缓存策略进行了研究。文献[4]基于智能域名解析设计了一种基站和核心网联合部署的 MEC 架构, 该架构实现了信息技术与通信技术的融合, 将上层功能下沉到网络边缘。文献[5]考虑用户访问资源的整体代价, 提出了一种基于效用的启发式分层协作缓存策略, 该策略根据资源的效用值进行决策, 可最小化用户资源

收稿日期: 2019-10-08; 修回日期: 2019-11-29; 录用日期: 2019-12-02。

基金项目: 国家自然科学基金资助项目(61173017); 国家重点研发计划(2018YFB2100300); 湖南省 2018 年移动互联网产业发展专项资金项目

作者简介: 柳兴(1984-), 男, 湖南岳阳人, 工程师, 博士, 主要研究方向: 移动边缘计算、移动云计算和 5G 行业应用; 杨震(1975-), 男, 湖南常德人, 副教授, 博士, 主要研究方向: 云计算、移动多媒体和 5G 行业应用; 王新军(1975-), 男, 湖南娄底人, 工程师, 硕士, 主要研究方向: 物联网、移动边缘计算和 5G 行业应用; 朱恒(1982-), 男, 湖南衡阳人, 工程师, 硕士, 主要研究方向: 移动边缘计算和 5G 行业应用。



访问代价。文献[6]考虑 MEC 服务器之间的协作,提出了一种协作式缓存管理算法,该算法可在最大化缓存命中率的同时最小化带宽开销。然而,上述方法都是从 MEC 协议或架构的角度进行设计,并没有考虑 MEC 的随机环境,仅从架构上来设计难以有效提升 MEC 的性能。

为了适应 MEC 的随机环境,进一步提升边缘计算的服务质量,文献[7]利用用户间的社交关系建立内容传播模型,采用传播动力学方法预测小用户数据下的内容流行度,从个体角度出发预测内容被访问的概率。文献[8]为了提升时延容忍任务的用户体验,通过获取移动用户的运动轨迹及个人偏好信息,设计了移动感知的服务调度算法。文献[9]基于移动用户位置信息提出一种区域人群流量预测的时空网络模型,通过融合各类外部特征信息,以短时局部流量降低对全局信息传输的要求。文献[10]针对视频应用场景,通过在边缘计算平台上执行视频分析服务操作,通过视频内容贴近用户放置来减少响应时间。虽然上述研究在很大程度上提升了移动边缘计算的性能,但是这些研究并没有关注 MEC 服务器存储空间的价值,因此还有一定的性能提升空间。

针对该问题,本文在考虑 MEC 服务器存储空间受限的基础上,利用移动用户群对不同对象的兴趣差异,提出了一种基于兴趣的内容分发加速策略(Interest-based Content Distribution Acceleration Strategy, ICDAS)。该策略根据 MEC 服务器存储空间、移动用户群对不同对象的兴趣以及对象的文件大小,选择性地在 MEC 服务器上缓存对象,最大限度的满足移动用户群的内容需求。仿真验证了所提策略的有效性。

1 系统模型

图 1 给出了 MEC 的网络架构。如图所示,MEC 服务器位于基站附近,可直接响应移动用户请求,也可将请求转发至后端的云数据中心。本文研究的内容分发加速问题,考虑 MEC 服务器存储空间受限、移动用户群对不同对象的兴趣差异,选择性地缓存对象数据。

假设 MEC 中的所有移动用户群的对象需求集合为 $\mathbb{A} = \{a_1, a_2, \dots, a_N\}$, MEC 服务器缓存任意对象 $a_i \in \mathbb{A}$ 占用的存储空间为 $p(a_i)$ 。假设 MEC 服务器的存储空间为 P ,且至少能够存储任意一个对象,则有 $\sum_{a_i \in \mathbb{A}} p(a_i) > P \geq \max_{a_i \in \mathbb{A}} p(a_i)$ 。

考虑移动用户群对内容分发加速的诉求是尽可能快地获取对象数据,同一对象数据被用户群体获取次数越多,则认为用户群对该对象的兴趣越高。考虑到 MEC 服务器存储空间受限,为了衡量用户群对单位存储空间内存储的对象内容的兴趣,引入兴趣度的概念。

定义 1 在 $[t_j, t_{j+1})$ 时段内,任意对象 $a_i \in \mathbb{A}$ 占用服务器的存储空间为 $p(a_i)$,被用户群访问的次数为 e_i ,则称该对象在 $[t_j, t_{j+1})$ 时段内的兴趣度为 $e_i/p(a_i)$ 。

本文的目标是根据不同时期移动用户群对不同对象的兴趣,设计一种缓存策略,提高 MEC 服务器的缓存命中率,提升用户感知。

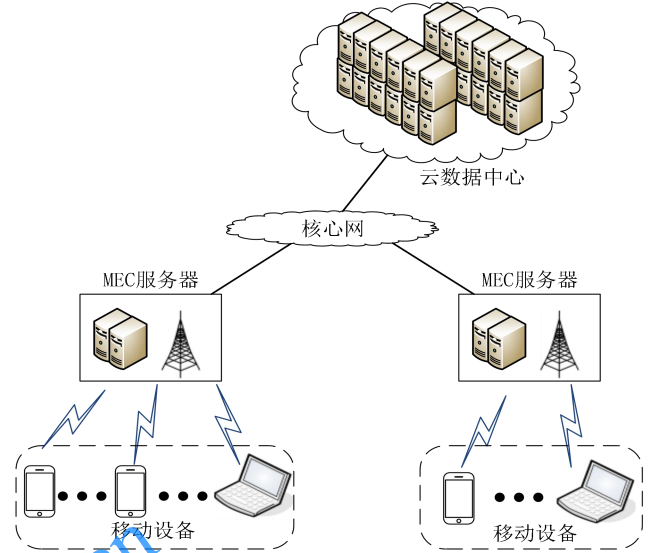


图 1 MEC 网络架构^[11]

Fig.1 Network frame of MEC^[11]

2 方案设计

为了减少 MEC 中的移动用户获取对象的延时,则 MEC 服务器缓存的对象必须具备一定的兴趣度。本文引入一个兴趣度排序的概念,利用这一概念能使 MEC 服务器尽量缓存兴趣度大的对象,提高 MEC 服务器的缓存命中率,相应的缓存特性用定理给出如下:

定理 1: 假设 MEC 服务器存储空间为 P , MEC 中的所有移动用户在 $[t_j, t_{j+1})$ 时段对集合 \mathbb{A} 的对象按兴趣度递减排序为 $\{a'_1, a'_2, \dots, a'_N\}$, 存在 $M < N$ 满足:

$$\sum_{1 \leq i \leq M, M < N} p(a'_i) \leq P \quad (1)$$

证明: 假设 MEC 服务器存储空间为 P , MEC 中的所有移动用户的对象需求集合为 $\mathbb{A} = \{a_1, a_2, \dots, a_N\}$, 则有:

$$\sum_{a_i \in \mathbb{A}} p(a_i) > P \quad (2)$$

假设所有移动用户在 $[t_j, t_{j+1})$ 时段对集合 \mathbb{A} 的对象按兴趣度递减排序为 $\{a'_1, a'_2, \dots, a'_N\}$ 。MEC 服务器至少能缓存任意一个对象,则有:



$$p(a'_i) \leq P \quad (3)$$

在 $[t_j, t_{j+1})$ 时段，移动用户感兴趣的对象一定，即

$$\sum_{a_i \in A} p(a_i) = \sum_{a_i \in A} p(a'_i), \text{ 则有:}$$

$$\sum_{a_i \in A} p(a'_i) > P \quad (4)$$

根据式 (3) 和 (4)，可推得存在 M ，满足 $M < N$ ，使得式 (1) 成立。证毕。

由定理 1 可知，MEC 服务器根据对象兴趣度排序便能极大地提升缓存命中率。由于 MEC 服务器是分布式架构，服务器会根据用户的需求做出响应，同时根据对象的兴趣度进行缓存。本文主要考虑被动拉的方式在 MEC 服务器上缓存用户感兴趣的对象，缓存控制的基本流程如图 2 所示。

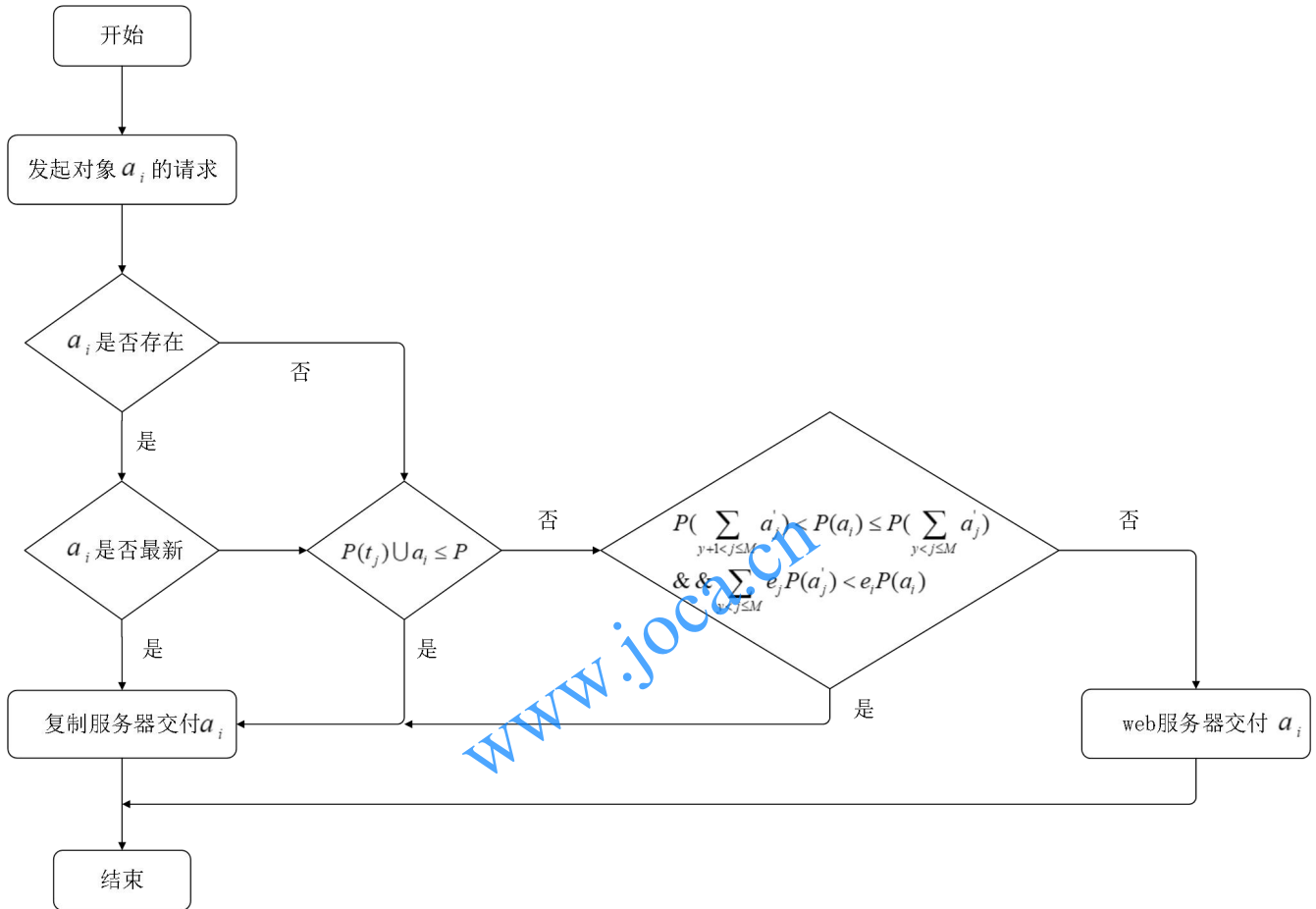


图 2 ICDAS 策略流程图

Fig.2 Flow diagram of ICDAS strategy

缓存控制策略的具体过程如下：

步骤一： MEC 服务器接收用户发起的对象 a_i 的请求，并判断是否存储对象 a_i 。若 MEC 服务器缓存了对象 a_i ，则跳至步骤二，否则跳至步骤三；

步骤二： 判断 MEC 服务器缓存的对象 a_i 是否为最新。若对象 a_i 为最新，则跳至步骤五，否则跳至步骤三；

步骤三： 判断 MEC 服务器是否有缓存空间存储对象 a_i 最新版本，即 $P(t_j) + P(a_i(t_j)) \leq P$ 是否成立。若成立则跳至步骤五，否则跳至步骤四；

步骤四： 判断 MEC 服务器用对象 a_i 替换其他对象，是否可减少云数据中心的访问量，即

$P(\sum_{1<j\leq y-1} a'_j) < P(a_i) \leq P(\sum_{y<j\leq M} a'_j) \&\& \sum_{y<j\leq M} e_j P(a'_j) < e_i P(a_i)$ 。若成立则跳至步骤五，否则跳至步骤六；

步骤五： MEC 服务器删除对象 $\{a'_j | y < j < M\}$ ，缓存并交付对象 a_i ，跳至步骤七；

步骤六： 云数据中心交付对象 a_i ，跳至步骤七；

步骤七： 控制策略结束。

该策略中， $P(t_j)$ 表示 MEC 服务器在 t_j 时刻缓存的数据量， $P(a_i(t_j))$ 表示对象 a_i 在 t_j 时刻的版本需要占用的缓存空间。由于策略采用用户拉的方式来缓存并更新 MEC



服务器的对象,因此在系统启动时,MEC 服务器不会缓存任何对象的数据。

3 仿真实验

仿真中,考虑移动用户向最近的 MEC 服务器发起申请,MEC 服务器与后端云数据中心同步数据的场景。假设总共有 1 亿个对象,所有对象的大小固定为 $p(a_i)=10$ MB,蜂窝小区为 1000 台移动终端在线,MEC 服务器存储空间 $P=100$ TB。本文从缓存命中率^[12]和延时两方面对所提策略的性能进行分析,并将所提策略与主动推送策略和周期同步策略进行比较。

图 3 给出了三种策略在缓存命中率方面的比较。从图中可以看出:(1)周期同步策略的缓存命中率会周期性的达到最大值,然后随着时间的推移递减;(2)主动推送策略的缓存命中率围绕某一值波动,验证了主动推送策略是将最新的对象推送给 MEC 服务器缓存,但是最新的对象不一定是用户最感兴趣的对象;(3)所提策略的缓存命中率随着仿真时间的推移而增加并逐渐趋于一个稳定值,这是因为所提策略是采用被动拉的方式进行缓存;(4)当所提策略的缓存命中率达到稳定后,所提策略要优于主动推送策略和周期同步策略,验证了采用移动用户对对象的兴趣度进行缓存的方式,可有效地提升缓存命中率。

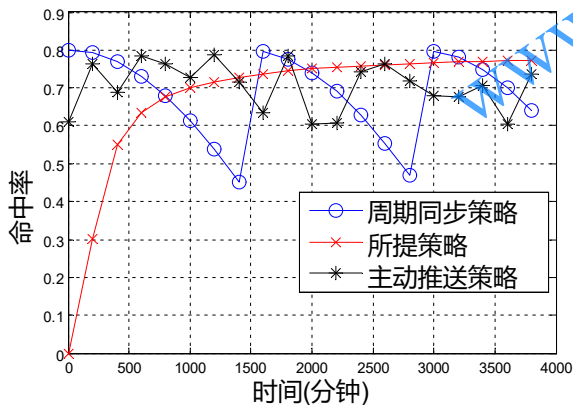


图3 缓存命中率比较

Fig.3 The comparison cache hit rates

图 4 给出了平均延时与用户数之间的关系。从图中可以看出:(1)随着用户数的增加,三种策略的延时都有所增加,这是因为 MEC 服务器的存储空间有限,用户数越大则对对象的种类需求大,必然导致部分用户到后端云数据中心取数据;(2)主动推策略优于周期同步策略,这主要是主动推策略能将最新的对象及时缓存至 MEC 服务器;(3)所提策略优于主动推策略,这是因为所提策略能根据小区用户兴趣在 MEC 服务器缓存对象。

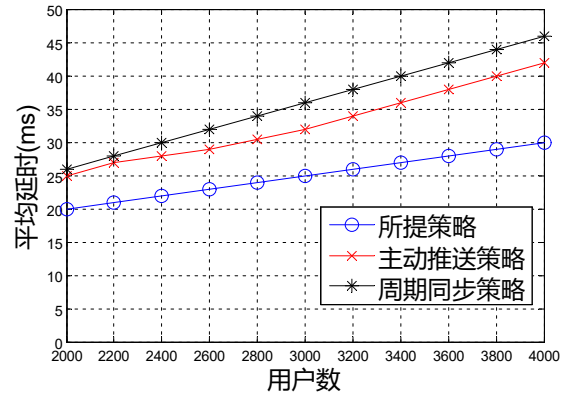


图4 平均延时与用户数的关系

Fig.4 The relation of average delay to user number

图 5 给出了平均延时与小区数之间的关系。从图中可以看出:(1)三种策略的平均延时随小区数的增加而增加,验证了不同小区用户感兴趣的对象不同,导致缓存命中率下降造成的结果;(2)所提策略的平均延时相对较平稳,且明显低于另外两种策略,这主要是因为所提策略充分考虑了不同小区用户需求的差异,能根据不同小区用户的需求,尽量缓存兴趣度高的对象;(3)结合图 4 和图 5 的仿真比较可得出结论,当系统运行达到稳定后,所提策略较另外两种策略,用户获取对象数据的时延可减少 20%。

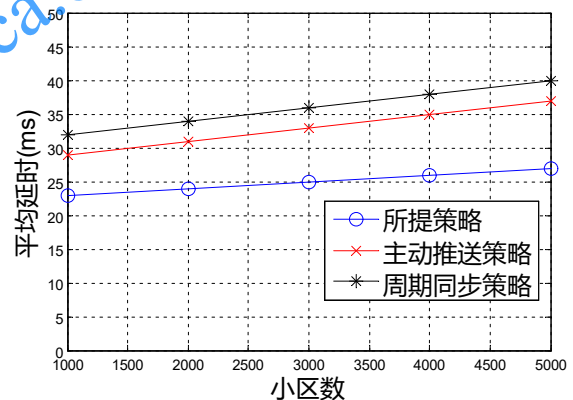


图5 平均延时与小区数的关系

Fig.5 The relation of average delay to cell number

4 结语

本文对 MEC 中的内容分发加速问题进行了研究,提出了一种 ICDAS 策略,该策略可根据不同小区用户对不同对象的兴趣的不同,进行差异化缓存,能够较好的适应 MEC 服务器存储受限的场景。仿真结果表明,所提策略在缓存命中率和延时方面均优于现有策略,在复杂的 MEC 环境中具有较好的应用前景。

文中所考虑的应用场景具有一定的合理性和有效性,在某种程度上可以提升移动边缘计算的性能。然而,在实际应用中,移动用户通常在一个区域内活动,相邻 MEC 服务器上缓存的对象存在一定的关系。下一步研究工作的重点是,



针对这种相邻 MEC 服务器缓存的对象存在一定关系的情况设计联合缓存, 以提高策略的实用性。

本文矢量、向量或矩阵。

参考文献

- [1] OUGHTON E, FRIAS Z, RUSSELL T, et al. Towards 5G: scenario-based assessment of the future supply and demand for mobile telecommunications infrastructure[J]. Technological Forecasting and Social Change, 2018:S0040162517313525.
- [2] MARKAKIS E K, KARRAS K, SIDERIS A, et al. Computing, caching, and communication at the edge: the cornerstone for building a versatile 5G ecosystem [J]. IEEE Communications Magazine, 2017, 55(11): 152-157.
- [3] 田辉, 范绍帅, 吕昕晨, 等. 面向 5G 需求的移动边缘计算[J]. 北京邮电大学学报, 2017, 40(2):1-10. (TIAN H, FAN S S, LV X C, et al. Mobile Edge Computing for 5G Requirements [J]. Journal of Beijing university of posts and telecommunications, 2017, 40(2):1-10.)
- [4] 陈昕, 温向明, 王鲁晗, 等. 5G 中多接入边缘计算的联合部署架构设计[J]. 北京邮电大学学报, 2018, 41(5):90-95. (CHEN X, WEN X M, WANG L H, et al. The Architecture Design of Cooperated Deployment for Multi-Access Edge Computing in 5G. Journal of Beijing university of posts and telecommunications, 2018, 41(5):90-95.)
- [5] 葛志诚, 徐恪, 陈靓, 等. 一种移动内容分发网络的分层协同缓存机制[J]. 计算机学报, 2018, 41(12):135-152. (GE Z C, XU G, CHEN L, et al. A Hierarchical Cooperative Caching Strategy for Mobile Content Delivery Network [J]. Chinese Journal of Computers, 2018, 41(12):135-152.)
- [6] BORST S, GUPTA V, WALID A. Distributed caching algorithms for content distribution networks[C]// Conference on Information Communications. IEEE Press, 2010:1478-1486.
- [7] SHUO H, HUI T, XINCHEN L. Edge popularity prediction based on social-driven propagation dynamics[J]. IEEE Communications Letters, 2017, 21(5):1-4.
- [8] CHAISIRI S, LEE B S, NIYATO D. Optimization of resource provisioning cost in cloud computing[J]. IEEE Transactions on Services Computing, 2012, 5(2):164-177.
- [9] 胡铮, 袁浩, 朱新宁, 等. 面向 5G 需求的人群流量预测模型研究[J]. 通信学报, 2019, 40(2):1-10. (HU Z, YUAN H, ZHU X, et al. Research on crowd flows prediction model for 5G demand[J]. Chinese Journal of Communications, 2019, 40(2):1-10.)
- [10] MAKINEN O. Streaming at the Edge: Local Service Concepts Utilizing Mobile Edge Computing[C]. //International Conference on Next Generation Mobile Applications, Services and Technologies. IEEE, 2016:1-6.
- [11] TANAKA H, YOSHIDA M, MPRI K, et al. Multi-access edge computing: A survey[J]. Journal of Information Processing, 2018, 26:87-97.
- [12] AHLEHAGH H, DEY S. Video-Aware Scheduling and Caching in the Radio Access Network[J]. IEEE/ACM Transactions on Networking, 2014, 22(5):1444-1462.

This work is partially supported by the National Natural Science Foundation of China (61173017), National Key Research and Development Program of China (2018YFB2100300) and Mobile Internet Development Special Foundation of Hunan Province.

LIU Xing, born in 1984, Ph. D., engineer. His research interests include mobile edge computing, mobile cloud computing and 5G industry applications.

YANG zhen, born in 1975, Ph. D., associate professor. His research interests include mobile cloud computing, mobile multimedia and 5G industry applications.

WANG Xinjun, born in 1975, M. S., engineer. His research interests include internet of things, mobile edge computing and 5G industry applications.

LIU Xing, born in 1984, M. S., engineer. His research interests include mobile edge computing and 5G industry applications.