

# 移动边缘计算中数据缓存和计算迁移的智能优化技术

## Intelligent Optimization of Data Caching and Computation Offloading in Mobile Edge Computing System

汪海霞/WANG Haixia  
赵志峰/ZHAO Zhifeng  
张宏纲/ZHANG Honggang

(浙江大学智能通信与网络安全研究所,  
杭州 310007)  
(Institute of Intelligent Communication  
and Network Security, Zhejiang  
University, Hangzhou 310007, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2018) 02-0015-004

**摘要:** 针对移动边缘计算(MEC),设计了计算迁移和数据缓存的联合优化模型,并基于改进的遗传算法(GA)对该模型的时延优化特性进行了优化。仿真表明:提出的方案比传统方案更能有效地降低用户请求的时延,对移动边缘网络的部署和应用具有一定的参考意义。

**关键词:** 移动边缘计算; 计算迁移; 数据缓存; 遗传算法

**Abstract:** In this paper, a combined optimization model for computing migration and data caching is designed. Particularly, an improved genetic algorithm (GA) is put forward to analyze the time delay optimization of the proposed model, which surely has certain significant values to the deployment and applications of mobile edge network.

**Keywords:** mobile edge computing; computation offloading; data caching; genetic algorithm

随着移动互联网和物联网(IoT)的快速发展以及各种新型业务(如虚拟现实(VR)、增强现实(AR)和视频会议等)的不断涌现<sup>[1]</sup>,用户对网络服务质量(QoS)的要求越来越高。因此,为了有效解决移动互联网发展带来的高负荷、低时延等要求的挑战,移动边缘计算(MEC)概念得以提出,并得到了学术界和产业界的广泛支持,被认为是下一代网络的关键技术之一<sup>[2-6]</sup>。

通过将相关的计算任务和请求数据迁移到近端(本地)MEC服务器上,可以减少网络设备能耗和传输时延,并大大提高用户体验。一般来说,计算迁移的关键技术就是充分利用MEC的优点(比如:缩短执行时延,降低能耗等),为此业务卸载过程决定合适的决策过程。文献[7]的工

作表明:将MEC与云计算相结合可以减少迁移任务的相关时延。文献[8]中,作者认为可以通过最小化任务的执行时延来实现最佳的迁移决策。在文献[9]中,作者提出了一种通过最小化能耗来进行决策的方法,其中优化问题被表述为一定约束条件下的马尔可夫决策过程(MDP)。此外,文献[10]和[11]分别分析了移动终端的能量消耗与多用户系统的相关时延间的关系。在文献[12]中,作者考虑在单个宏小区的微基站(BS)上进行数据缓存,通过最小化宏基站服务的请求数量来优化缓存方案。文献[13-18]中,作者也都研究分析了基站缓存中的存储分配问题。

虽然MEC具有较强的计算和数

据传输能力,但相对于现在急速增加的移动终端数量与业务,MEC的相关资源仍然十分有限。本文基于MEC与数据中心(DC)的相互配合机制,根据排队理论分析了每个任务的平均执行时延<sup>[9]</sup>,然后提出了一定缓存空间约束下的时延最小化问题,并通过改进的遗传算法解决该优化问题,从而有效降低了用户请求时延,提高了缓存性能和效率。

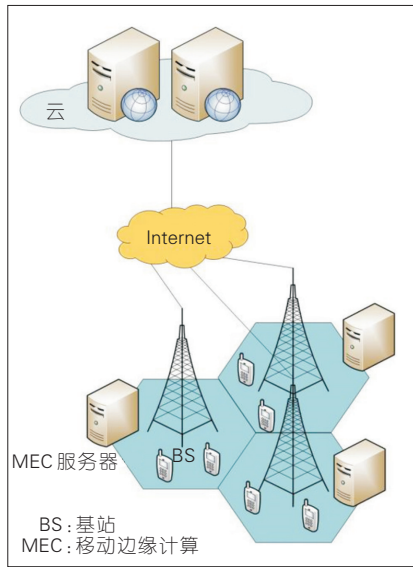
### 1 系统模型

如图1所示,MEC系统有 $B$ 个BS(每个BS配备有一个MEC服务器)和 $M$ 个用户设备(UE)。在这里我们可以把MEC服务器看作是一个具有计算和存储能力的小型数据中心,通过

收稿日期: 2018-01-12

网络出版日期: 2018-03-17

基金项目: 中兴通讯研究项目“基于认知SDN的5G-RAN技术研究”资金资助



▲图1 MEC系统

部署与边缘交换机关联的虚拟机,将相关的计算任务迁移在 MEC 服务器上,也可以将用户所请求的内容数据缓存在 MEC 服务器上。假设来自 UE 的业务请求的到达过程遵循泊松分布,则 UE  $i$  的请求速率被表示为  $\lambda_i$ 。 $c_i$  为 UE  $i$  所请求的计算任务,  $d_i$  表示 UE  $i$  所要访问的数据量,  $e_j$  表示 MEC 服务器  $j$  的数据缓存容量。把 BS 覆盖范围内每个 UE 当做 M / M / 1 排队系统进行处理,则 MEC 服务  $j$  和 UE  $i$  的服务速率分别可以定义为  $\mu_j$  和  $\theta_j$ 。

通常情况下,移动终端发送请求,然后由 BS 接受处理或将请求发送往远端云端进行处理,这个过程涉及 BS 和云,以及 UE 和 BS 之间的传输。为了简单起见,将 BS 与云之间传输的单位时延表示成  $e_{MC}$ ,将 UE 与 BS 传输过程中的单位时延表示为  $e_{UM}$ 。用变量  $x_{ij}$  表示服务器  $j$  是否缓存了用户  $i$  所请求的内容,如果服务器内缓存了用户请求的内容,则变量  $x_{ij}$  为 1,否则为 0。同样地,如果将用户  $i$  产生的计算任务迁移到 MEC 服务器  $j$  上进行处理,变量  $y_{ij}$  为 1;反之,若进行本地处理,则为 0。

根据排队论<sup>[9]</sup>,我们可以计算得出服务器  $j$  处理计算任务产生的时

延,即:

$$C_j^M = y_{ij} c_i \lambda_i / (\mu_j - \sum_{i=1}^M y_{ij} \lambda_i) \quad (1)$$

其中,  $T_i^{UM} = \sum_{j \in B} e_{UM} (y_{ij} c_i + d_i)$  是迁移到服务器的总计算任务量,注意  $\mu_j - \sum_{i=1}^M y_{ij} \lambda_i > 0$ 。

同样,由用户  $i$  进行本地处理相关任务的时延是:

$$C_i^U = (1 - \sum_{j \in B} y_{ij}) c_i \lambda_i / (\theta_i - (1 - \sum_{j \in B} y_{ij}) \lambda_i) \quad (2)$$

其中,  $(1 - \sum_{j \in B} y_{ij}) c_i \lambda_i$  是由用户本地处理的计算任务量,并需要满足  $\theta_i - (1 - \sum_{j \in B} y_{ij}) \lambda_i > 0$ 。

以上是对用户产生的计算任务,在不同的处理方式下计算的执行时延。对于在数据传输过程中产生的传输时延,如果 MEC 服务器缓存了用户所请求的数据,则用户和服务器之间的传输时延表示为  $T_i^{UM} = \sum_{j \in B} e_{UM} (y_{ij} c_i + d_i)$ ;若服务器缓存中没有用户所请求的内容数据,则服务器和数据中心之间传输数据所得的时延为  $T_i^{MC} = \sum_{j \in B} e_{MC} (1 - x_{ij}) d_i$ ,其中  $1 - x_{ij} = 0$  意味着用户所请求的内容缓存在服务器里,而无需访问数据中心。

综上所述,所有用户产生的总时延为请求的执行时延和传输时延的总和,即:

$$T = \sum_{i=1}^M C_j^M + C_i^U + T_i^{UM} + T_i^{MC} \quad (3)$$

其优化问题就可以表述如下:

$$\text{Minimize } T$$

$$\text{s.t. } \sum_{i=1}^M d_i x_{ij} \leq \varepsilon_j, \quad \forall j \in B, \forall i \in M \quad (4)$$

$$\mu_j - \sum_{i=1}^M y_{ij} \lambda_i > 0, \quad \forall j \in B, \forall i \in M \quad (5)$$

$$\theta_i - (1 - \sum_{j \in B} y_{ij}) \lambda_i > 0, \quad \forall j \in B, \forall i \in M \quad (6)$$

$$x_{ij} \in \{0, 1\} \quad \forall j \in B, \forall i \in M \quad (7)$$

$$y_{ij} \in \{0, 1\} \quad \forall j \in B, \forall i \in M \quad (8)$$

其中,约束条件(4)限制了缓存

在服务器的数据总量不应该超过服务器的总容量。其他的约束条件上文都有提到,不再赘述。

## 2 改进的遗传算法

遗传算法(GA)<sup>[20]</sup>是模拟自然界生物进化过程与机制求解优化问题的一类自适应的启发式智能搜索算法。该算法通过将初始个体进行选择、交叉和变异等操作来更新种群。其最主要的特征是种群搜索策略和种群个体之间的信息交换,非常适用于处理传统方法不容易解决的非线性以及复杂的问题,其应用领域非常广。模拟退火算法(SA)<sup>[21]</sup>是一种贪心算法,它在搜索过程中加入了随机因子,并且以一定的概率来接受比较差的解,这样就有可能避免陷入局部最优。

为了更加有效地解决上述优化问题,与传统的遗传算法相比,我们提出了如下改进遗传算法(算法1):

(1)在交叉过程中,对新的种群进行了模拟退火操作,使得一些适应值较低的个体也有一定的概率遗传到下一代,这样可以有效防止算法收敛到局部最优,并使算法更加稳定。

(2)在种群进行交叉操作和变异操作之后,分别计算新种群里面个体的适应度函数。这样就可以防止一些优秀的个体在种群进化过程中被遗弃掉。

在改进的算法1中,种群个体是指上述优化问题中的变量  $\{x_{ij}, y_{ij}\}$ ,即前  $M$  维的值为  $x_{ij}$  的取值,后  $M$  维的值为  $y_{ij}$  的取值,所以个体是一个用 0 和 1 填充的  $2 \times M$  维的向量。值得注意的是:在所改进的算法中给出的可行解都满足约束方程(4)、(5)和(6),适应度函数就是所要优化的目标函数。

### 算法1 改进的遗传算法

初始化种群 Pop;  
根据公式(3)计算适应度;

将 Pop 中的最优个体进行存储;  
 While ( $i <$  最大迭代次数) do  
   根据交叉概率对种群 Pop 进行交叉得到新种群 Pop1;  
   计算种群 Pop1 的适应度;  
   并将种群 Pop1 中的最优个体与初始化存储的进行比较, 保存最优的个体;  
   根据变异概率对种群 Pop1 进行变异得到新种群 Pop2;  
   计算种群 Pop2 适应度并进行比较, 保存最优个体;  
   用模拟退火算法更新种群 Pop2;  
   计算适应度并保存最优的个体;  
 End  
 Return 最优个体以及对应的适应度函数

### 3 仿真设计与分析

为了评估本文提出的基于改进遗传算法的联合计算和缓存优化方法, 我们将其与其他两种策略进行了性能对比: 一种是不考虑边缘的计算和存储资源的传统策略, 即本地执行所有的计算任务并且所有请求的数据都存储在数据中心; 另一种是随机策略, 它虽然考虑了移动边缘网络的相关资源, 但采用随机方式来决定是在终端还是在服务器上执行计算任务, 并且数据也是随机存储在边缘服务器或数据中心。随机策略在一定程度上牺牲了服务质量, 从而降低了数据缓存过程的复杂性。

在仿真过程中, 除非文中明确说明, 否则所有主要参数值均基于表 1 进行设置。

图 2 描述了当用户数  $M = 200$  时, 在不同策略下总时延随着服务器缓存容量 (从 500 增加到 3 500) 的变化关系。可以观察到传统策略完全不受服务器缓存容量的变化, 当缓存容量从 500 增加到 1 000 时, 我们提出的策略的总时延从 510.2 ms 急剧下降到 320.1 ms。这是由于随着服务器的缓存容量的增加, 通过改进的遗传

算法的自适应调节使得数据传输时间变短。但当容量达到一定的阈值之后, 由于用户请求数量是一定的, 所以总时延减少的较为缓慢。

图 3 描述了当用户数  $M$  从 50 增加到 400 时, 不同策略的总时延变化情况。我们可以观察到: 与其他策略相比, 我们所提出的策略实现了较低的延迟, 并且可以减少超过 50% 的执行延迟。特别是当用户数量很少时,

边缘服务器有足够的资源用较少的时间来处理大部分任务。随机策略利用了边缘网络的计算和存储资源, 因此其性能优于传统策略。另外, 不同策略下的总时延间的差距也随着用户数目的增加而增大, 意味着当 MEC 系统的规模较大时, 所提出的策略具有较大的性能提升。这是由于提出的策略可以根据用户请求数据大小的不同来提前缓存, 从而使得数

▼表 1 仿真参数设置

参数	值	参数	值
$M$	[50,500]	$B$	[1,5]
$d_i$	[1,20]/MB	$c_i$	[1,5]/MB
$\lambda_i$	[1,10]/s	$\mu_j$	[1.5,2] $\times 10^3$
$\theta_j$	[0.5,1] $\times 10^3$	$e_{MC}$	1.5 s/MB
$e_{UM}$	1 s/MB	$\varepsilon_j$	1 000/MB

图 2  
不同策略下总时延与 MEC 服务器缓存容量的关系

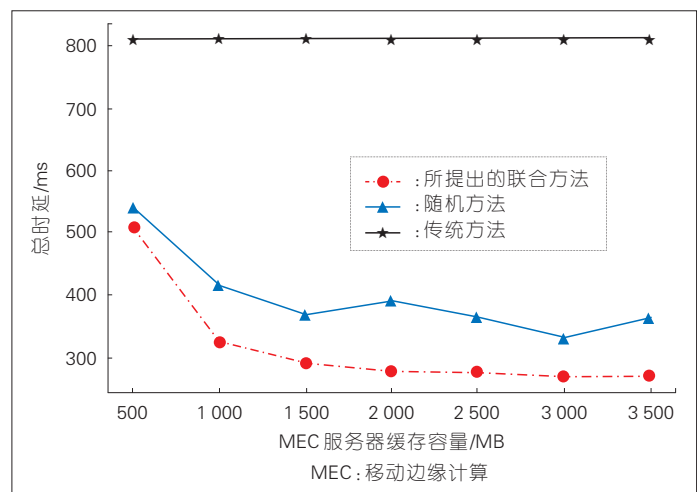
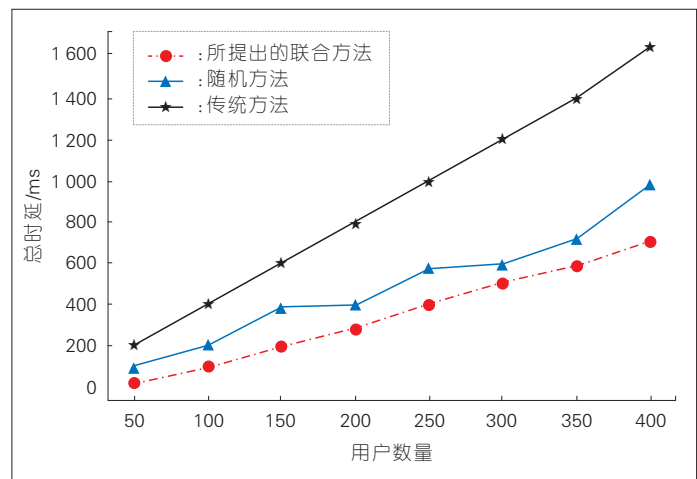


图 3  
不同策略下总时延与用户数的关系



据的传输时延较低。

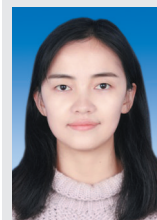
## 4 结束语

在 MEC 中,计算迁移和数据缓存是很重要的特性,缓存决策策略的优劣直接影响移动边缘计算系统的性能。本文中我们通过联合分析执行时延和传输时延,用改进的遗传算法来建立优化的计算迁移和数据缓存模型,有效提高了缓存空间的使用效率,性能方面也有较大程度的提升。

### 参考文献

- [1] MACH P, BECVAR Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading [J]. IEEE Communications Surveys & Tutorials, 2017, (99): 1-1
- [2] WANG X, 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. DOI: 10.1109/MCOM.2014.6736753
- [3] European Telecommunications Standards Institute (ETSI). Mobile Edge Computing Introductory Technical White Paper[R], 2014
- [4] HU Y C, PATEL M, SABELLA D, et al. Mobile Edge Computing: A Key Technology towards 5G[J]. ETSI White Paper, 2015,11(11):1-16
- [5] KUMAR K, LIU J, LU Y H, et al. A Survey of Computation Offloading for Mobile Systems [J]. Mobile Networks and Applications, 2013, 18(1):129-140
- [6] 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. DOI: 10.1109/MCOM.2014.6871674
- [7] RIMAL B P, VAN D P, MAIER M. Mobile-Edge Computing vs. Centralized Cloud Computing in Fiber-Wireless Access Networks[C]//2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). USA:IEEE,2016: 991-996
- [8] LIU J, MAO Y, ZHANG J, et al. Delay-Optimal Computation Task Scheduling for Mobile-Edge Computing Systems[C]//2016 IEEE International Symposium on Information Theory (ISIT).USA:IEEE, 2016: 1451-1455. DOI: 10.1109/ISIT.2016.7541539
- [9] KAMOUN M, LABIDI W, SARKISS M. Joint Resource Allocation and Offloading Strategies in Cloud Enabled Cellular Networks [C]//2015 IEEE International Conference on Communications (ICC).USA:IEEE, 2015: 5529-5534
- [10] JIANG Z, MAO S. Energy Delay Tradeoff in Cloud Offloading for Multi-Core Mobile Devices[J].IEEE Access, 2015, 3:2306-2316. DOI: 10.1109/ACCESS.2015.2499300
- [11] KWAK J, KIM Y, LEE J, et al. Dream: Dynamic Resource and Task Allocation for Energy Minimization in Mobile Cloud Systems[J].IEEE Journal on Selected Areas in Communications, 2015, 33(12):2510-2523. DOI: 10.1109/JSAC.2015.2478718
- [12] POULARAKIS K, IOSIFIDIS G, TASSIULAS L. Approximation Algorithms for Mobile Data Caching in Small Cell Networks[J].IEEE Transactions on Communications, 2014, 62(10):3665-3677. DOI: 10.1109/TCOMM.2014.2351796
- [13] GU J, WANG W, HUANG A, et al. Proactive Storage at Caching-Enable Base Stations in Cellular Networks[C]//2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC). USA:IEEE, 2013:1543-1547. DOI: 10.1109/PIMRC.2013.6666387
- [14] BASTU E, BENNIS M, KOUNTOURIS M, et al. Cache-Enabled Small Cell Networks: Modeling and Tradeoffs[C]// Wireless Communications Systems (ISWCS), 2014 11th International Symposium on.USA:IEEE, 2015. DOI: 10.1109/ISWCS.2014.6933434
- [15] BLASCO P, GUNDUZ D. Learning-Based Optimization of Cache Content in A Small Cell Base Station[C]//2014 IEEE International Conference on Communications (ICC). USA:IEEE, 2014: 1897-1903. DOI: 10.1109/ICC.2014.6883600
- [16] GOLREZAEI N, SHANMUGAM K, DIMAKIS A G, et al. Femtocaching: Wireless Video Content Delivery through Distributed Caching Helpers[C]// Proceedings of IEEE INFOCOM 2012.USA:IEEE, 2012: 1107-1115. DOI: 10.1109/TIT.2013.2281606
- [17] POULARAKIS K, IOSIFIDIS G, SOURLAS V, et al. Multicast-aware Caching for Small Cell Networks[C]//2014 IEEE Wireless Communications and Networking Conference (WCNC). USA:IEEE, 2014:2300-2305. DOI: 10.1109/WCNC.2014.6952688
- [18] POULARAKIS K, IOSIFIDIS G, TASSIULAS L. Approximation Caching and Routing Algorithms for Massive Mobile Data Delivery [C]//2013 IEEE Global Communications Conference (GLOBECOM).USA: IEEE, 2013. DOI: 10.1109/GLOCOM.2013.6831621
- [19] COOPER R B. Introduction to Queueing Theory[M]. New York: North Holland, 1981
- [20] DAVIS L. Handbook of Genetic Algorithms [R]. Van Nostrand Reinhol, 1991
- [21] KIRKPATRICK S, GELATT C D, VECCHI M P, et al. Optimization by Simulated Annealing [J]. Science, 1983, 220(4598):671-680

### 作者简介



**汪海霞**, 浙江大学信息与电子工程学院硕士研究生在读; 主要研究方向为移动边缘计算、智能数据缓存技术。



**赵志峰**, 浙江大学信息与电子工程学院副教授; 主要研究方向为认知无线电、无线 mesh 网络、移动边缘计算和 SDN 在无线通信中的应用; 已发表论文 100 余篇, 申请发明专利 20 余项。



**张宏纲**, 浙江大学信息与电子工程学院教授; 主要研究方向为移动通信、人工智能、机器学习、认知无线电、绿色通信和下一代异构蜂窝网络架构; 已发表论文 200 余篇, 提出 IEEE 802.15 国际标准提案达 16 项, 并拥有 3 项国际专利。