



5G边云协同分布式网络架构下任务卸载模型

陈慧敏, 胡玉佩

(广东邮电职业技术学院, 广东 广州 510630)

【摘要】 自动驾驶、增强现实等5G新兴的应用对网络时延和可靠性提出更高的挑战, 针对边缘协同框架负载均衡问题, 提出一种大规模边云协同分布式网络架构下的任务卸载模型。该模型以最低能耗为目标, 根据任务特性与现有网络资源、计算资源和存储资源自适应优化任务卸载决策, 通过整合边缘计算与云计算处理能力的优势, 保证时延敏感型任务的质量, 提高整个系统的负载均衡, 降低能耗。

【关键词】 边云协同; 任务卸载; 资源管理; 负载均衡

doi:10.3969/j.issn.1006-1010.2021.04.024 中图分类号: TN929.5

文献标志码: A 文章编号: 1006-1010(2021)04-0144-05

引用格式: 陈慧敏. 5G边云协同分布式网络架构下任务卸载模型[J]. 移动通信, 2021,45(4): 144-148.

OSID:



扫描二维码
与作者交流

Task Offloading Model Based on 5G Edge Cloud Collaborative Distributed Network Architecture

CHEN Huimin, HU Yupei

(Guangdong Vocational College of Post and Telecom, Guangzhou 510630, China)

[Abstract] 5G emerging applications such as autopilot and augmented reality pose higher challenges to network delay and reliability. Aiming at the load balancing problem of edge collaboration framework, this paper proposes a task offloading model under large-scale edge cloud collaborative distributed network architecture. The model takes the minimum energy consumption as the goal, and adaptively optimizes the task unloading decision according to the task characteristics and the existing network resources, computing resources and storage resources. By integrating the advantages of edge computing and cloud computing, the delay sensitive task quality is guaranteed, the load balance of the whole system is improved, and the energy consumption is reduced.

[Keywords] edge cloud collaboration; task unloading; resource management; load balancing

0 引言

5G 场景下大规模边云协同分布式网络架构需要应对系统资源负载均衡的问题, 云计算系统通常采用网络资源和计算资源联合优化的方式来实现系统的负载均衡。但 5G 边云协同分布式网络架构相比传统的云计算系统具有独特的特征: 随机变化的无线网络参数、网络资源和计算资源容易出现时空耦合现象。这些特征为网络资源和计算资源联合优化带来新的挑战。

为了解决上述问题, 相关学者结合网络资源和计算资源的动态变化状态, 采用任务卸载的方式实现系统资源负载均衡, 比如: 任晨珊^[1]提出李雅普诺夫方法分散

数据处理压力, 该方法结合网络拓扑、链路链条与服务 器处理能力来设计虚拟队列偏置和计算负载均衡, 并通过边缘服务器和云端中心的卸载来应对突发性、密集型的数据量; 郭俊^[2]针对超密集网络架构中移动边缘计算存在的问题, 提出联合前后向链路计算策略及分布式任务调度方案为用户接入和计算卸载提供更多选择性; 郝哲^[3]以最小化 MEC 延迟和能量消耗为优化目标, 提出一种以用户选择并行优化迭代计算卸载、无线频谱资源调度和计算资源调度的联合优化策略; Fajardo 等人^[4]采用调度基站边缘侧服务器的服务级别来调整 LTE 下行业务流的延迟; Liu 等人^[5]结合应用程序缓冲区状态、移动终端用户、MEC 服务器计算能力以及终端与 MEC 服务器信道特征来决定是否将应用卸载到 MEC 服务器上, 以实现最小化

收稿日期: 2020-12-14

执行延迟的目标；Barbarossa 等人^[6]提出移动终端的计算卸载策略，该策略在满足最大执行约束限制下最小化移动用户终端的能量消耗，最终降低用户等待队列长度。

部分学者对移动终端的计算卸载策略产生一定的质疑，原因在于：在计算卸载过程中，将数据传送到基站边缘服务器所消耗的能量明显低于移动终端处理数据所消耗的能量。上述的学者认为，计算卸载应该围绕用户移动场景下对延迟和消耗进行权衡取舍。随着 5G 应用的激增，系统中卸载应用程序所需要的时间就会越来越多，基站侧边缘服务器处理业务所需要的时间将会越来越久，因此，当用户将终端应用卸载到基站侧边缘服务器后，由于边缘服务器处理能力的限制，将会引发一系列新的问题：如何从边缘服务器资源现状和负载均衡目标出发，提出边缘和云端服务器间的资源调度与任务卸载方案；如何根据业务 QoS 特性、节点负载状态自适应优化任务卸载决策等。

为了解决上述的问题，本文提出一种大规模边云协同分布式网络架构下的任务卸载模型，该模型以最低能耗目标，在保证移动任务的服务质量下，以高效能的方式实现网络资源、计算资源和存储资源的合理调度和任务执行，快速响应计算请求，提高整个系统的负载均衡。

1 大规模边云协同分布式网络架构

大规模边云协同分布式网络架构包括三个部分，主要有边缘云、核心网云以及控制云。整个架构将不同计

算和存储能力的服务器分散到整个网络中，利用网络功能虚拟化技术和软件定义网络技术实现计算和虚拟资源的动态调度，实现全网资源的弹性化管理。大规模边云协同分布式网络架构如图 1 所示。

边缘云是部署到基站侧的网络服务实施层，手机和电脑设备通过无线链路接入基站侧的边缘云，边缘云借助有限的计算能力将数据进行预处理并将结果传至核心网，进行业务集中处理。边缘节点计算能力和带宽资源差异性较大，而且边缘资源会随着业务请求的变化而动态变化，边缘云控制器一旦发现边缘云处理能力不足，为减少等待队列长度，边缘服务器将会将数据卸载到核心网云。

核心网云由多个云计算中心网络构成，拥有丰富的存储资源和强大的数据处理能力，可以保存边缘云卸载的数据以及执行复杂任务的能力。核心网云利用网络功能虚拟化技术和软件定义网络技术对网络的各种资源进行管理、调度和优化，以应对海量业务请求。除此之外，核心网云借助动态负载均衡技术实现资源池的动态扩展和缩减，在保证业务性能前提下降低网络传输和计算开销。因此，核心网云处理数据的性能在速度、成本以及可扩展性方面优于边缘云。

控制云是基于局部和全局开销进行资源调度和管理，其通过软件定义技术实现网络资源使用效率的监测。控制云会根据边缘云和核心网云各自负载情况和网络状态决定资源是否卸载，以高能效的方式快速响应业务请求，

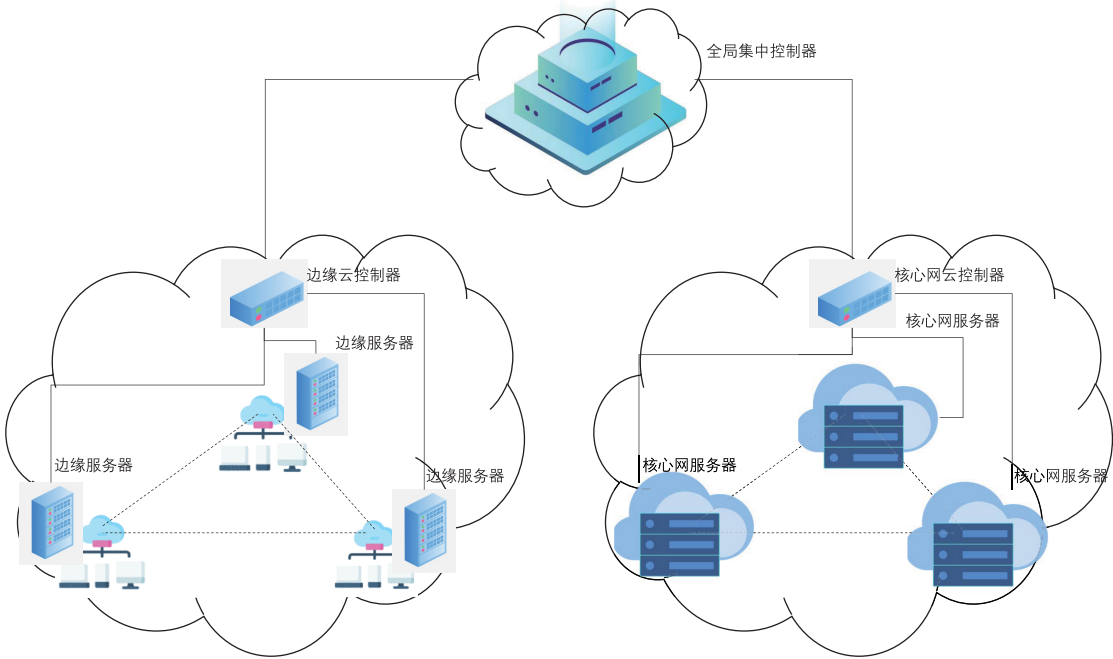


图1 大规模边云协同分布式网络架构

满足 5G 业务差异性的需求。

大规模边云协同分布式网络架构利用了边缘云架构和核心网云架构的各自优势来应对密集任务请求，分散网络处理数据的压力，保证整个网络的正常运行。

2 边云协同分布式网络架构的任务卸载策略

2.1 边云协同分布式网络架构的任务卸载模型

任务卸载可以有效解决边云协同分布式网络架构在资源存储、计算性能以及任务等待队列不均衡存在的问题，它通过卸载决策决定哪些计算需要放在边缘服务器处理，哪些计算需要卸载到核心网服务器处理，最后结合服务器现有开销情况和任务时延确定任务卸载位置。本文方案主要考虑用户请求的 QoS 需求以及边缘计算资源等要素，寻找能耗最小化，实现全网性能最大化。

假设 $W = \{w_1, w_2, \dots, w_i, \dots, w_n\}$ 表示一组需要完成的任务，其中 $w_i(d_i, l_i)$ 表示子任务 i 所需要的计算开销和数据量。子任务通过边缘计算节点或者核心网服务器处理后返回到用户，由于子任务选择不同节点执行所产生的能耗和传输时延是不一样的，因此，本文分别计算任务在两种情况处理的能耗和时延。

(1) 边缘节点执行的能耗时延卸载模型

如果任务 i 在边缘节点执行时，此时终端能耗主要由数据从终端发送能耗和终端空闲能耗两部分组成。那么终端能耗计算公式为：

$$E_c = P_t \times \frac{l_i}{R_{MRC}} + T_{MEC} \times P_w \quad (1)$$

其中， P_t 为终端数据发送的功率， l_i 为任务 i 数据量大小， R_{MRC} 为终端发送给边缘节点数据速率， P_w 为终端空闲的功率。 T_c 为任务的处理时间，由边缘服务器的计算能力决定。由于数据传输的时延（微秒级别）通常小于数据处理（毫秒级别）的时延，因此本文的时延仅仅考虑数据处理的时延 T_c 。 T_{MEC} 如式 (2) 所示：

$$T_{MEC} = \sum_{i=1}^n \frac{d_i}{f_m} \quad (2)$$

其中， d_i 为任务 i 计算量大小， f_m 表示边缘计算节点的计算能力。

边缘计算节点的能耗是指执行任务 i 所引起的能耗，而影响节点能耗有 CPU 的利用率、内存利用率以及硬盘利用率，其中硬盘利用率采用 I/O 利用率来表征。

边缘计算节点的功率估算公式参考文献 [7]，可以表示为：

$$P_{MEC}(t) = w_{cpu} \times U_{cpu}(t) + w_{mem} \times U_{mem}(t) + w_{io} \times U_{io}(t) + \epsilon \quad (3)$$

其中， w_{cpu} 、 w_{mem} 、 w_{io} 是 CPU、内存以及硬盘的非线性相关系数，由边缘节点的具体配置和节点使用情况决定，而 $U_{cpu}(t)$ 、 $U_{mem}(t)$ 、 $U_{io}(t)$ 是边缘计算节点 CPU、内存以及硬盘在 t 时刻的利用率，不同时刻的利用率差异性很大。 ϵ 是一个误差补偿系数。

边缘计算节点的功率估算公式可以表示为：

$$E_{MEC} = T_{MEC} \times P_{MEC} \quad (4)$$

那么，边缘节点执行的能耗时延卸载计算公式为：

$$E_M = E_c + E_{MEC} = P_t \times \frac{l_i}{R} + T_{MEC} \times (P_{MEC} + P_w) \quad (5)$$

(2) 云端执行的能耗时延卸载模型

任务 i 在云端执行时，能耗主要是终端发送能耗、终端空闲能耗以及云端服务器处理的能耗。那么终端能耗计算公式为：

$$E_c = P_t \times \frac{l_i}{R_{cloud}} + T_{cloud} \times P_w \quad (6)$$

其中， P_t 为终端数据发送的功率， l_i 为任务 i 数据量大小， R_{cloud} 为终端发送给云端节点数据速率， P_w 为终端空闲的功率。 T_{cloud} 为任务的处理时间，由云端服务器的计算能力决定。由于数据传输的时延远远小于数据处理的时延相比，因此本文的时延仅仅考虑数据处理的时延 T_{cloud} 。

$$T_{cloud} = \sum_{i=1}^n \frac{d_i}{f_m} \quad (7)$$

其中， d_i 为任务 i 计算量大小， f_m 表示云端计算节点的计算能力。

云端计算节点的能耗是指执行任务 i 所引起的能耗，而影响节点能耗有 CPU 的利用率、内存利用率以及硬盘利用率，其中硬盘利用率采用 I/O 利用率来表征。

云端计算节点的功率估算公式可以表示为：

$$P_{cloud}(t) = w'_{cpu} \times U'_{cpu}(t) + w'_{mem} \times U'_{mem}(t) + w'_{io} \times U'_{io}(t) + \epsilon' \quad (8)$$

其中， w'_{cpu} 、 w'_{mem} 、 w'_{io} 是 CPU、内存以及硬盘的线性非相关系数，由云端节点的具体配置和节点使用情况决定，而 $U'_{cpu}(t)$ 、 $U'_{mem}(t)$ 、 $U'_{io}(t)$ 是云端节点 CPU、内存以及硬盘在 t 时刻的利用率，不同时刻的利用率差异性很大。 ϵ' 是一个误差补偿系数。

云端计算节点的功率估算公式可以表示为：

$$E_{cloud} = T_{cloud} \times P_{cloud} \quad (9)$$

那么，边缘节点执行的能耗时延卸载计算公式为：

$$E_C = E_c + E_{cloud} = P_t \times \frac{l_i}{R} + T_{cloud} \times (P_{cloud} + P_w) \quad (10)$$

2.2 任务卸载策略

以最低能耗为目标，综合业务时延、业务 CPU 性能、存储性能、带宽性能和用户 QoS 等指标，判定业务是否从边缘节点卸载到云端节点，选择合适的节点进行数据

处理。卸载策略的公式表达为:

$$\begin{aligned} & \min \{E_C, E_M\} \\ & s.t. \\ & V_i \geq V \\ & T \geq T_i \\ & c_i + \sum_{j \in V} c_j \leq C_i \\ & m_i + \sum_{j \in V} m_j \leq M_i \\ & b_i + \sum_{j \in V} b_j \leq B_i \end{aligned} \quad (11)$$

其中, 限制项的第一部分 V_i 表示任务 i 传输速率, V 表示该类任务的最小传输速率。限制项的第二部分 T_i 表示任务 i 时延, V 表示该类任务的最小时延。限制项的第三部分表示将某一个任务 i 部署到物理节点 i 中, 那么已经占用的计算资源和该任务 i 所需要的计算资源之和要小于物理节点 i 最大的计算资源。限制项的第四部分表示将某一个任务 i 部署到物理节点 i 中, 那么已经占用的存储资源和该任务 i 所需要的存储资源之和要小于物理节点 i 最大的存储资源。限制项的第五部分表示将某一个任务 i 部署到物理节点 i 中, 那么已经占用的带宽资源和该任务 i 所需要的带宽之和要小于物理节点 i 最大带宽资源。

3 实验分析

本文采用仿真方式对能耗最优的任务卸载策略进行验证。整个实验平台由 16 台云服务器组成, 其中边缘计算云由 12 台服务器组成, 核心网云由 4 台服务器组成。每台服务器的配置如下: 四核 2.8 GHz 的 CPU 处理器, 每核有 2 个线程; 内存为 8 G; 硬盘为 1 T 以及配有千兆的以太网接口。为了验证本文的算法, 本文在随机时间段设置海量视频识别密集处理任务, 观察当计算量和带宽量都很大的时候, 基于随机卸载的算法和基于本文的卸载策略的算法在节能方面的影响。所谓节能表示不采用任务卸载策略相比能量节省比例。

随机任务卸载是指边缘节点一旦发现自身处理能力不足时, 将任务随机卸载到满足性能需求的核心网云的服务器中, 这种卸载方式没有从全局考虑核心网云计算系统中资源的占用情况, 仅仅考虑服务器性能和业务的 QoS 性能需求。

本文任务卸载算法是指边缘节点一旦发现自身处理能力不足时, 对核心网云计算系统中资源的占用情况、服务器性能以及业务的 QoS 性能需求将任务卸载到能耗使用最低的服务器中。

实验的第一部分是不考虑时延要求的任务量对边云协同处理节能的影响, 如图 2 所示:

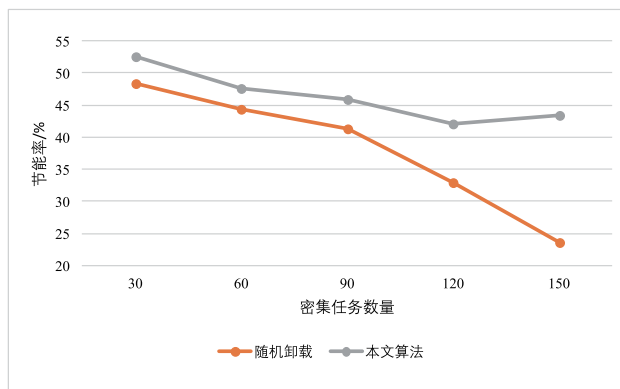


图2 密集任务数量对不同卸载算法的节能影响 (不考虑时延)

如图 2 所示, 当密集任务量较少的时候, 两种卸载算法的节能效果都较为趋同, 一旦任务量迅速提升, 而且相关节点已经被占用的情况下, 系统的节能效果开始下降。随机卸载算法节能效果迅速下降, 而采用本文算法的节能可达到 40% 左右。

实验的第二部分是不考虑时延要求的任务量对边云协同处理节能的影响。当单位任务处理的时延要求很高时, 不同的卸载算法在系统节能方面的影响如图 3 所示:

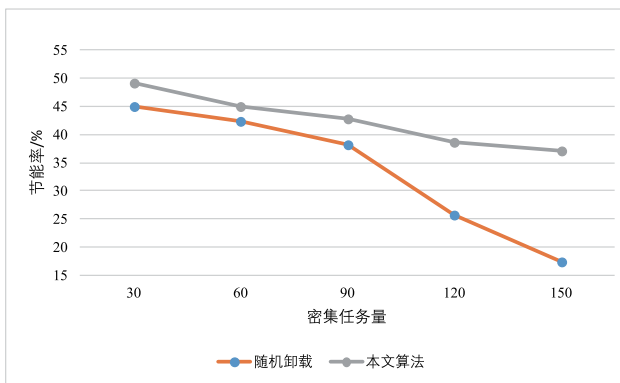


图3 密集任务数量对不同卸载算法的节能影响 (考虑时延)

由图 3 可知, 当单任务处理时间的时延要求较高, 采用本文卸载方法在节能方面的效果更好, 这是因为本文从系统计算节点实际使用情况和任务 QoS 需求出发, 采用云边协同的方式对资源进行合理分配, 将一些计算量较大的任务卸载到云端, 不仅利用了云端的超强处理能力降低数据处理的时延, 还采用资源协同优化方式保证系统的负载均衡, 从而使得系统整体的能耗降低。

4 结束语

本文研究边云协同分布式网络架构下采用不同的任务卸载策略在节能方面的不同表现。实验证明, 本文以最低能耗为目标, 综合服务器 CPU 性能、存储性能、带

宽性能和用户 QoS 等指标, 判定业务采用哪种卸载策略以达到最优的节能效果, 该策略能够在缺少未来业务负载预测的情况下优化节点资源管理和调度, 动态适应实时动态变换的工作负载和业务多变的性能要求, 最终以高效低能耗的方式完成 5G 网络海量数据的处理。

参考文献:

- [1] 任晨珊. 边缘计算中的高效与负载均衡技术研究 [D]. 北京: 北京邮电大学, 2019.
- [2] 郭俊. 超密集网络中基于移动边缘计算的卸载策略研究 [D]. 北京: 北京邮电大学, 2018.
- [3] 郝哲. 移动边缘计算的计算卸载和资源调度策略研究 [D]. 北京: 北京工业大学, 2019.
- [4] Fajardo J O, Taboada I, Liberal F. Radio-aware service-level scheduling to minimize downlink traffic delay through mobile edge computing[C]//International Conference on Mobile Networks and Management. Springer. Cham, 2015: 121-134.
- [5] 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). IEEE, 2016: 1451-1455.
- [6] Barbarossa S, Sardellitti S, Di Lorenzo P. Joint allocation of computation and communication resources in multiuser mobile cloud computing[C]//2013 IEEE 14th workshop on signal processing advances in wireless communications (SPAWC). IEEE, 2013: 26-30.
- [7] 卢彪, 李悦, 张万礼. 基于云计算系统使用率的能耗模型的研究与设计 [J]. 绥化学院学报, 2016,36(9): 148-151.
- [8] 王文礼, 贾芝婷. 边缘计算与云协同问题研究 [J]. 河北省科学院学报, 2020,37(4): 9-14.
- [9] Ren J, Yu G, He Y, et al. Collaborative cloud and edge computing for latency minimization[J]. IEEE Transactions on Vehicular Technology, 2019,68(5): 5031-5044.
- [10] Qi Q, Tao F. A smart manufacturing service system based on edge computing, fog computing, and cloud computing[J]. IEEE Access, 2019(7): 86769-86777.
- [11] Wang K, Yin H, Quan W, et al. Enabling collaborative edge computing for software defined vehicular networks[J]. IEEE Network, 2018,32(5): 112-117.

作者简介



陈慧敏 (orcid.org/0000-0002-9991-1044): 讲师, 硕士毕业于华中科技大学, 现任职于广东邮电职业技术学院信息通信工程学院, 研究方向为通信网及交换技术。

胡玉佩: 讲师, 硕士毕业于北京邮电大学, 现任职于广东邮电职业技术学院, 主要研究方向为移动通信技术、数据通信技术。

(上接第119页)

道状态估计和在有限探索的基础上, 在业务 QoS 约束下以系统吞吐量最大化为目标实现可用信道的选择, 会在很大程度上提升了终端侧的接入成功率和系统平均吞吐量。实验表明, 本文无线资源管理策略不仅能够节省信道资源开销, 还能提升频谱利用的效率, 提升用户满意度。

参考文献:

- [1] Huang J, Yin Y, Qiang D, et al. A Game-Theoretic Analysis on Context-Aware Resource Allocation for Device-to-Device Communications in Cloud-Centric Internet of Things[C]//International Conference on Future Internet of Things & Cloud. IEEE Computer Society. IEEE, 2015:80-85.
- [2] Zhang G, Yang K, Liu P, et al. Joint Channel Bandwidth and Power Allocation Game for Selfish Cooperative Relaying Networks[J]. IEEE Transactions on Vehicular Technology, 2012,61(9): 4142-4156.
- [3] Tian Y, Liu H, Li Z. Power Allocation Algorithm Based on Non-cooperative Game Theory for LTE Downlink[C]//IEEE/WIC/ACM International Conferences on Web Intelligence & Intelligent Agent Technology. IEEE, 2012: 643-647.
- [4] 刘海燕. 面向 5G 的超密集网络中分布式无线资源管理的研究 [D]. 北京: 北京交通大学, 2016.
- [5] 张月莹. 基于 QoE 的无线资源管理算法研究 [D]. 北京: 北京邮电大学, 2013.
- [6] 杨宁. 基于深度学习的无线资源管理与安全研究 [D]. 北京: 北京科技大学, 2020.
- [7] 陈慕涵, 郭佳佳, 李潇, 等. 基于深度学习的大规模 MIMO 信道状态信息反馈 [J]. 物联网学报, 2020(1): 33-44.
- [8] 汪周飞, 袁伟娜. 基于深度学习的多载波系统信道估计与检测 [J]. 浙江大学学报 (工学版), 2020,54(4): 732-738.
- [9] 陈嘉润, 余宝贤, 王剑莹, 等. 基于深度神经网络的高频谱效率频分复用系统的信道估计方法 [J]. 华南师范大学学报: 自然科学版, 2020,52(3): 22-26.
- [10] Qin Z, Ye H, Li G Y, et al. Deep Learning in Physical Layer Communications[J]. IEEE Wireless Communications, 2019: 1-7.

作者简介



陈慧敏 (orcid.org/0000-0002-9991-1044): 讲师, 硕士毕业于华中科技大学, 现任职于广东邮电职业技术学院信息通信工程学院, 研究方向为通信网及交换技术。