

**本科毕业论文（** **设计** **）** **中期报告**

**题** **目：面向公平性的云边微服务系统部署方法**

**专** **业** **软件工程**

**学** **生** **付书煜**

**学** **号**  **2021111824**

**指导教师** **贺祥**

**日** **期**  **2025 年**  **3 月** **1 日**

**哈尔滨工业大学教务处制**

**目** **录**

[1 论文（设计）工作是否按开题报告预定的内容及进度安排进行 1](#bookmark3)

[2 已完成的研究工作及成果 2](#bookmark4)

2.1 已完成的研究工作 2

2.2 成果 2

[3 后期拟完成的研究工作及进度安排 3](#bookmark6)

[3.1 视频流服务的云-边协同架构的系统建模 4](#bookmark8)

[3.2 响应时间差异最小化的公平性优化算法 4](#bookmark10)

[3.2.1 优化目标及约束条件 4](#bookmark12)

[3.2.2 优化算法设计概述 4](#bookmark14)

[3.3 模型公平性优化效果的评估标准 4](#bookmark16)

[3.3.1 评估指标 5](#bookmark18)

[3.3.2 评估方法 5](#bookmark20)

[4 存在的问题与困难 5](#bookmark22)

[4.1 视频流服务的云-边协同架构的系统建模 5](#bookmark24)

[4.2 公平性优化算法设计 7](#bookmark26)

[4.2.1 优化目标的具体建模 7](#bookmark28)

[4.2.2 优化算法的实现 8](#bookmark30)

[4.3 模型公平性优化效果的评估方法 9](#bookmark32)

[5 论文按时完成的可能性 9](#bookmark34)

1 论文（设计）工作是否按开题报告预定的内容及进度安排进行

随着互联网的迅速发展，微服务架构因其高效、灵活的特性被广泛应用于视频流媒体、在线游戏等对实时性和灵活性要求较高的场景中。微服务通过将单体应用拆分为独立、细粒度的模块，使各模块能够独立部署和扩展，并通过轻量级通信协议进行协作，从而大幅提升了系统的灵活性和扩展性。然而，这种架构也带来了网络延迟、系统管理复杂、资源利用效率低等问题，尤其是在高并发访问和复杂任务处理时，如何高效分配和管理资源成为亟待解决的关键问题。

为应对这些挑战，云-边协同计算模式应运而生。该模式通过在云端和边缘节点之间合理分配任务，最大化各自的资源优势。云端资源丰富，适合处理大规模数据和复杂任务，但由于其物理距离远，网络传输延迟较高；边缘节点则位于用户附近，能够就近处理请求，显著降低延迟从而提升用户体验。然而，边缘节点资源有限，易在高并发或复杂任务处理中出现性能瓶颈。因此，在云-边协同环境中，不仅要合理分配任务和资源，还需根据计算需求和网络条件，灵活调整服务实例的部署，以更好地发挥云端和边缘节点的资源优势，从而提升整体 QoE（服务质量体验）。

现有的云-边协同部署方案多集中于最大化整体 QoE，但在优化整体用户体验时，未充分考虑不同用户之间的体验差异和公平性问题。由于地理位置、网络条件、节点资源负载的差异，不同用户的响应时间波动较大，尤其是远离边缘节点的用户，可能因网络延迟较高，体验显著低于接近边缘节点的用户。对于视频流媒体等对实时性和一致性要求较高的应用场景，这种 QoE 不均衡会严重影响用户体验。

公平性问题可以从两个方面来考虑。首先，不同优先级的用户在服务响应上存在显著差异。例如，在视频流媒体平台上，购买 VIP 会员的高优先级用户期望能够以更高的画质、更快的加载速度观看视频，边缘节点可以优先处理这些高优先级用户的请求，从而减少加载时间和缓冲现象；而普通用户则主要关注视频的流畅播放，不希望出现长时间的加载或低画质的问题。高优先级用户的期望和实际体验之间的差距，可能会影响他们对平台的满意度。其次，即使在同一优先级的用户之间，网络条件和设备性能差异也可能导致体验上的不公平。例如，尽管同为 VIP 会员的用户，他们的网络条件不同，可能会造成体验上的巨大差异。网络较差的高优先级用户可能在观看视频时出现缓冲或延迟现象，而设备较旧的普通用户可能会面临视频卡顿或加载缓慢的问题，这些因素都可能导致同一优先级用户之间的体验差异。对于视频流媒体等对实时性和一致性要求极高的应用场景，这种 QoE 不均衡会严重影响用户的体验。

因此，如何通过优化服务实例部署和资源分配，减少不同优先级之间的响应时间差异并均衡同一优先级用户的 QoE，已成为云-边协同计算中亟待解决的重要问题。

近年来，国内外学者围绕云-边协同计算中的 QoE 优化和公平性展开广泛研究。早期为提升整体 QoE，提出基于用户位置、任务分配和缓存策略的方法，像边缘节点就近处理、云边协同调度，虽能降低延迟、提升 QoE，却侧重单一层次性能优化，未充分考虑用户间 QoE 差异与公平性。后续部分研究关注公平性优化，运用动态资源分配、负载均衡、深度强化学习等方法减少 QoE 差异，例如用图神经网络进行动态工作负载迁移，用深度强化学习优化任务卸载和缓存策略，但这些方法在大规模分布式环境下计算效率和调度机制有待优化。目前，大多数研究集中于单一层次资源调度，缺乏对云-边协同环境中微服务部署的整体考量，未能有效解决 QoE 不均衡和用户公平性问题，如基于区块链的资源分配虽提升了透明性和公平性，却存在多资源类型综合调度和平衡不完善、远离边缘节点用户 QoE 差异仍存在的情况。

综上，如何结合云端和边缘资源，借助智能化手段对部署策略加以优化，提升公平性，增强体验一致性，正是本研究致力于攻克的核心问题。工作的进度安排如下：

|  |  |  |
| --- | --- | --- |
| 工作安排 | 周数 | 起止时间 |
| 系统建模与设计阶段 | 3 | 2024.11.20-2024.12.15 |
| 算法设计与实现 | 5 | 2024.12.16-2025.01.19 |
| 仿真实验与数据收集 | 4 | 2025.01.20-2025.02.16 |
| 敏感性分析与模型调优 | 3 | 2025.02.17-2025.03.09 |
| 成本效益分析与优化验证 | 4 | 2025.03.10-2025.04.06 |
| 结题，论文编写 | 2 | 2025.04.07-2025.04.20 |

目前为止，基本按照开题报告中的规划进行，已完成了系统建模、算法设计以及部分仿真实验。

2 已完成的研究工作及成果

2.1 已完成的研究工作--面向公平性的云边微服务系统部署方法

如开题报告中的场景描述：作为一家提供智能网络服务的公司，我们负责为一个大规模的用户群体设计和优化其服务架构。用户分布在不同的地理位置，既有处于城市中心的用户，也有偏远地区的用户，他们通过不同的终端设备连接到我们的服务系统。为了确保服务质量，特别是在响应时间上的公平性，我们引入了云边协同架构，在各地部署边缘节点，同时将计算资源丰富的云服务器与边缘节点紧密结合。

边缘节点由于距离用户较近，能够减少物理传输延迟，但由于其资源有限，可能在负载过高时导致响应时间变慢。而云服务器虽然提供更强大的计算能力和资源，但由于距离用户较远，可能导致更高的传输延迟。因此，我们需要合理调度用户的请求，决定是将其分配给边缘节点还是云服务器。

在这个场景中，用户的优先级（例如高付费用户和普通用户）会影响资源的分配，优先级较高的用户会获得更多的计算资源，从而保证更低的响应时间。响应时间不仅受到服务器资源的影响，还与数据的传输延迟和处理能力密切相关。此外，边缘节点和云服务器的部署成本和资源负载也需要考虑在内，以避免过度依赖边缘节点导致的超载现象。

我们的目标是通过优化资源分配和服务实例的部署，使得不同优先级的用户都能够在合理的响应时间内获得服务，尽可能提高系统的公平性。我们还需要确保云边协同架构在部署实例时能够平衡成本和资源消耗，避免不必要的资源浪费或过度负载，从而实现系统的高效运行和经济可行性。

2.1.1 云-边协同架构的系统设计与建模

我们假设有个用户，每个用户有一定的数据需求，表示该用户请求的视频流数据量。用户之间的地理位置不同，可能存在不同的网络延迟。用户的位置用二维坐标表示。用户的优先级用表示，用户优先级越高，值越大。此外，我们有个服务器，表示所有可供用户连接的服务器集合，包括边缘服务器集和云服务器集。边缘服务器通常部署在接近用户的区域，以提供低延迟服务，其位置也用二维坐标表示，分布在用户覆盖范围的内部或周边区域。相比之下，云服务器位置通常固定在远离用户的中心化数据中心，以提供大规模计算和存储资源。每个服务器上可能部署了一个或多个服务实例；如果没有用户连接到某个服务器，该服务器可以不部署任何服务实例，处于“非活跃”状态。服务实例的数量受到节点资源限制的约束。服务器和用户之间的连接通过变量来表示，为1表示用户连接到服务器，为0则表示没有连接。用户连接到到服务器时，服务器需要分配一定的资源来处理其视频流请求。资源需求集合表示了这些资源的具体量。其中、和分别表示用户所需的CPU、内存和带宽资源量。用户和服务器之间的物理距离可通过用户和服务器的二维坐标计算得到，如公式（4-1）所示。

(4-1)

用户请求从到服务节点的延迟由传输延迟和处理延迟组成。

对于边缘节点，传输延迟包括物理传输延迟和带宽延迟，物理传输延迟可以通过用户到边缘服务节点的距离和信号传播速度来计算，如公式（4-2）；而带宽延迟则与用户请求数据大小以及边缘网络带宽有关，如公式（4-3）。边缘节点总传输延迟为公式（4-4）所示。边缘节点的处理延迟由其处理能力决定，假设边缘节点的处理速率为，则其处理延迟为公式（4-5）所示。

(4-2)

(4-3)

(4-4)

(4-5)

同样的，云节点的传输延迟和处理延迟也可按照类似的方式计算，但其云节点通常具有更高的带宽和处理能力。云节点的延迟计算公式如（4-6）、（4-7）所示。

(4-6)

(4-7)

式中，为云节点的物理传输延迟，云节点的带宽延迟，为云节点的处理速率。

综合响应时间由传输延迟和处理延迟组成。若用户连接到边缘节点，则响应时间为公式（4-8），若用户连接到云节点，则响应时间为公式（4-9）：

(4-8)

(4-9)

在上述响应时间计算的基础上，引入加权响应时间的概念，用以进一步体现不同用户优先级对响应时间的影响。加权响应时间定义为用户的响应时间与其优先级的乘积，如（4-10）所示：

= (4-10)

对于服务实例的部署，边缘节点和云节点的资源不同，边缘节点通常资源有限，而云节点资源相对充足。每个节点部署成本的计算方式如下：

边缘节点的部署成本可以表达为公式（4-11），其中是固定成本，是基于用户需求所消耗的CPU、内存和带宽资源的成本,可以通过用户请求量和节点资源的消耗来计算，如公式（4-12）：

(4-11)

(4-12)

式中，、和为边缘节点的资源单价。

云节点部署成本可以表示为云资源使用成本和网络流量成本之和，如公式（4-13），其中云资源使用成本与边缘节点资源使用成本计算方法类似，如公式（4-14）；而网络流量成本是根据用户从不同区域访问云节点所产生的额外网络传输费用来计算的,如公式（4-15）：

(4-13)

(4-14)

(4-15)

式中，表示云平台的流量单价。

故总的部署成本可以表示为所有边缘节点和云节点的部署成本之和，如（4-16）：

(4-16)

2.2 成果

3 后期拟完成的研究工作及进度安排

4 存在的问题与困难

4.1 视频流服务的云-边协同架构系统建模

5 论文按时完成的可能性

5.1 进度安排

5.2 预期目标

（1）**中期目标** 完成云边协同环境中视频流服务的公平性优化模型设计和算法的基 础框架，进行初步实验与参数调优。

（2）**结题目标** 实现云边协同环境下的公平性优化模型，能够基于视频流服务的实 际需求，在满足资源和成本限制的前提下，最大限度地减少用户间响应时间差异。

6 课题已具备和所需的条件、经费

本课题的研究工作已经具备了充足的条件和经费保障。首先，研究将充分利用现有的 实验室资源和仿真平台，这些平台能够支持大规模的云-边协同架构实验，并进行多种性 能评估。实验室内拥有必要的硬件设备以及高性能计算资源，为研究提供了充分的技术保 障。与此同时，相关经费也已充足，能够覆盖研究过程中所需的软硬件支出、数据采集与 处理、人员支持等各项费用。因此，课题的顺利开展具有良好的资源基础和保障。

7 研究过程中可能遇到的困难和问题，解决的措施

（1）**数据真实性不足** 仿真数据可能无法完全模拟实际应用场景的数据特征，影响 实验结果的可靠性；解决措施：在仿真数据的基础上，可以通过公开数据集进行结果验证， 提高模型的适用性。

（2）**约束条件冲突** 多个约束条件（如延迟、成本、资源分配）之间可能存在冲突， 影响优化结果；解决措施：采用多目标优化方法，平衡不同目标之间的关系。

8 主要参考文献

[1] J. Thönes, "Microservices," in IEEE Software, vol. 32, no. 1, pp. 116-116, Jan.-Feb. 2015, doi: 10. 1109/MS.2015.11.

[2] Ashwin Rao, Arnaud Legout, Yeon-sup Lim, Don Towsley, Chadi Barakat, and Walid Dabbous. 2011. Network characteristics of video streaming traffic. In Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies (CoNEXT ' 11). Association for Computing Machinery, New York, NY, USA, Article 25, 1–12. <https://doi.org/10.1145/2079296.2079321>

[3] Dragoni, N. *et al.* (2017). Microservices: Yesterday, Today, and Tomorrow. In: Mazzara, M., Meyer, B. (eds) Present and Ulterior Software Engineering. Springer, Cham.

<https://doi.org/10.1007/978-3-319-67425-4_12>

[4] 赵然,朱小勇.微服务架构评述[J]. 网络新媒体技术,2019,8(1):58-61F0003

[5] 李春霞.微服务架构研究概述[J].软件导刊, 2019, 018(008):1-3,7.

[6] J. Pan and J. McElhannon, "Future Edge Cloud and Edge Computing for Internet of Things Applications," in IEEE Internet of Things Journal, vol. 5, no. 1, pp. 439-449, Feb.

2018,doi: 10. 1109/JIOT.2017.2767608.

[7] M. N. O. Sadiku, S. M. Musa and O. D. Momoh, "Cloud Computing: Opportunities and Challenges," in IEEE Potentials, vol. 33, no. 1, pp. 34-36, Jan.-Feb. 2014, doi: 10. 1109/MPOT.2013.2279684.

[8] W. Zhang, Y. Hu, Y. Zhang, and D. Raychaudhuri, “ SEGUE: Quality of service aware edge cloud service migration,” in Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.,

2016, pp. 344–351.

[9] Q. He et al., “A game-theoretical approach for mitigating edge DDoS attack,” IEEE Trans. Dependable Secure Comput., vol. 19, no. 4, pp. 2333–2348, Jul./Aug. 2022.

[10] K. Cao, Y. Liu, G. Meng and Q. Sun, "An Overview on Edge Computing Research," in IEEE Access, vol. 8, pp. 85714-85728, 2020, doi: 10. 1109/ACCESS.2020.2991734.

[11] X. He, H. Xu, X. Xu, Y. Chen and Z. Wang, "An Efficient Algorithm for Microservice Placement in Cloud-Edge Collaborative Computing Environment," in IEEE Transactions on Services Computing, vol. 17, no. 5, pp. 1983-1997, Sept.-Oct. 2024, doi: 10. 1109/TSC.2024.3399650.

[12] H. SHI, R. V. Prasad, E. Onur and I. G. M. M. Niemegeers, "Fairness in Wireless Networks:Issues, Measures and Challenges," in IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 5-24, First Quarter 2014, doi: 10. 1109/SURV.2013.050113.00015.

[13] P. Lai et al., “QoE-aware user allocation in edge computing systems with dynamic QoS,” Future Gener. Comput. Syst., vol. 112, pp. 684–694, 2020.

[14] F. Liu, G. Tang, Y. Li, Z. Cai, X. Zhang and T. Zhou, "A Survey on Edge Computing Systems and Tools," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1537-1562, Aug. 2019, doi: 10. 1109/JPROC.2019.2920341.

[15] R. Alsurdeh, R. N. Calheiros, K. M. Matawie and B. Javadi, "Hybrid Workflow Scheduling on Edge Cloud Computing Systems," in IEEE Access, vol. 9, pp. 134783-134799, 2021, doi: 10. 1109/ACCESS.2021.3116716.

[16] 马璐, 刘铭, 李超, 路兆铭, 马欢. 面向6G 边缘网络的云边协同计算任务调度算[J]. 北京邮电大学学报, 2020, 43(6): 66-73.

[17] 王朝, 高岭, 高全力, 等. 边缘计算中基于博弈论的数据协作缓存策略研究[J].计算 机应用研究, 2020, 37 (12): 3739-3743.

[18] 张雅洁,陆旭,李曦,等.电力物联网下基于云边协同的计算任务放置算法[J]. 电力信息 与通信技术,2024,22(10):38-47.DOI:10. 16543/j.2095-641x.electric.power.ict.2024.10.06.

[19] 朱 仪 , 江 雪 . 基 于 云 边 协 同 的 任 务 卸 载 策 略 技 术 研 究 [J/OL]. 无 线 电 工 程,1-21[2024-11-13].<http://kns.cnki.net/kcms/detail/13.1097.tn.20241010.0924.004.html>.

[20] 张 文康 ,赵伟 ,刘 德 成 , 等 . 云边 协 同低 延迟 故 障 预 测算 法研 究 [J]. 能源 与环 保,2024,46(10):238-243.DOI:10. 19389/j.cnki.1003-0506.2024.10.036.

[21] C. Zhang, J. Yin and S. Deng, "Ensuring Fairness in Edge Networks: A GNN-Based Media Workload Migration Scheme With Fairness Guarantee," in IEEE Transactions on Services Computing, vol. 17, no. 3, pp. 934-948, May-June 2024, doi: 10. 1109/TSC.2023.3298695.

[22] 肖 旋 . 云 边 协 同 场 景 下 服 务 器 负 载 均 衡 与 协 调 研 究 [D]. 重 庆 大 学,2021.DOI:10.27670/d.cnki.gcqdu.2021.004253.

[23] 冯起,薛喜红,任龙,等.考虑云端距离的科技服务边缘计算资源均衡调度算法[J]. 自动 化技术与应用,2024,43(08):95-98+104.DOI:10.20033/j.1003-7241.(2024)08-0095-05.

[24] H. Hao, C. Xu, W. Zhang, S. Yang and G. -M. Muntean, "Computing Offloading With Fairness Guarantee: A Deep Reinforcement Learning Method," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 10, pp. 6117-6130, Oct. 2023, doi: 10. 1109/TCSVT.2023.3255229.

[25] J. Zhou, F. Chen, Q. He, X. Xia, R. Wang and Y. Xiang, "Data Caching Optimization With Fairness in Mobile Edge Computing," in IEEE Transactions on Services Computing, vol. 16, no. 3, pp. 1750-1762, 1 May-June 2023, doi: 10. 1109/TSC.2022.3197881.

[26] 吴 忠 辉 . 基 于 区 块 链 的 边 缘 分 布 式 计 算 卸 载 关 键 技 术 研 究 [D]. 北 京 邮 电 大 学,2024.DOI:10.26969/d.cnki.gbydu.2024.000039.

[27] 徐恒 炜 . 基 于 公 平 通 信 方 案 的 穿 戴 设 备 隐 私 学 习 方 法 [D]. 中 国 科 学 技 术 大 学,2023.DOI:10.27517/d.cnki.gzkju.2023.002058.

[28] 张 世 焱 . 云 边 协 同 网 络 中 的 多 资 源 管 理 机 制 研 究 [D]. 北 京 邮 电 大 学,2024.DOI:10.26969/d.cnki.gbydu.2024.000220.

[29] 金韬,庄丽婉,张晨,等.基于区块链的云边协同系统研究与设计[J]. 信息安全研 究,2021,7(04):310-318.