

Web 前端与边缘计算协同的低延迟应用架构设计与实现

阿布里米提·肉孜

新疆师范高等专科学校（新疆教育学院），新疆 乌鲁木齐 830043

DOI:10.61369/EST.2024050002

摘 要： 目的：随着物联网与实时交互应用的兴起，传统云计算架构下的高延迟问题日益凸显，设计了一种 Web 前端与边缘计算协同的低延迟应用架构，方法：通过前端智能分流，边缘节点就近处理与动态资源调度三层协同机制，降低了应用响应时间。结果：架构在视频分析，游戏以及工业控制等场景下，相比传统云端部署可减少 58% 的延迟并提高 43% 的用户体验满意度，结论：Web 前端与边缘计算的深度融合能有效突破地理分布式应用的延迟瓶颈，为实时性要求高的 Web 应用提供了新的架构范式。

关 键 词： 边缘计算；Web 前端；低延迟架构；实时应用；计算卸载

Design and Implementation of Low-latency Application Architecture for the Collaboration of Web Front-End and Edge Computing

Abulimiti · Rouzi

Xinjiang Teacher's College (Xinjiang Education Institute), Urumqi, Xinjiang 830043

Abstract : Objective: With the rise of the Internet of Things and real-time interaction applications, the high latency problem under the traditional cloud computing architecture has become increasingly prominent. A low-latency application architecture featuring the collaboration of the Web front end and edge computing is designed. Method: Through a three-layer collaborative mechanism of intelligent front-end diversion, nearby processing of edge nodes, and dynamic resource scheduling, the application response time is reduced. Result: In scenarios such as video analysis, gaming, and industrial control, the architecture can reduce latency by 58% and increase user experience satisfaction by 43% compared to traditional cloud deployment. Conclusion: The deep integration of Web front-end and edge computing can effectively break through the latency bottleneck of geographically distributed applications, providing a new architectural paradigm for Web applications with high real-time requirements.

Keywords : edge computing; web front-end; low-latency architecture; real-time application; computational offloading

引言

随着 5G 网络与物联网技术的快速发展，以 AR/VR、工业互联网与智慧城市为代表的新兴应用对网络延迟提出了严苛要求，传统云计算架构将计算任务集中处理，导致地理上分散的用户面临较高的网络延迟，边缘计算通过将计算资源下沉至网络边缘，为解决这一问题提供了技术支撑，然而现有技术多关注服务端部署优化忽略了 Web 前端的积极作用，通过 Web 前端与边缘计算的协同设计可充分利用终端设备，边缘节点以及云中心的分层特性，构建真正端到端的低延迟应用架构，为延迟敏感型应用开辟新的发展空间。

一、Web 前端与边缘计算协同架构设计

（一）边缘计算与 Web 前端技术发展

边缘计算作为云计算的延伸与补充，通过将计算资源布置在网络边缘，能够减少数据传输距离降低网络延迟，目前主要的边缘计算平台包括 AWS Wavelength、Azure Edge Zones 与 Google Distributed Cloud Edge 等^[1]。边缘计算技术呈现出轻量级虚

拟化、开放标准以及边缘智能等发展趋势。

与此同时 Web 前端技术也飞速发展，从单纯的展示层发展为具备复杂计算能力的平台，主要技术进展包括 Web Worker、WebAssembly、Progressive Web Apps (PWA) 与 WebRTC 等^[2]。特别是 WebAssembly 技术的出现，使 Web 前端具备了处理计算密集型任务的能力，为前端与边缘计算协同提供了技术基础，WebAssembly 相比 JavaScript 在计算密集型任务

上性能提升显著，处理速度平均提高3-6倍。

（二）低延迟应用需求与挑战

不同类型的应用对延迟的敏感度各不相同，典型应用场景如VR/AR交互要求延迟低于20ms，在线游戏需要低于50ms，工业控制甚至要求延迟低于10ms，这些严苛的延迟要求给传统应用架构带来了巨大挑战，主要受到地理距离，网络拥塞以及处理延迟等因素的限制，传统云计算架构难以满足上述场景的低延迟需求，迫切需要新型架构来应对这些挑战^[3]。Web前端与边缘计算协同架构正是为解决这些问题而设计的，它能够利用终端设备的计算能力，就近边缘节点的处理能力与云服务的强大资源，形成多层协同的处理模式。

（三）协同架构总体设计

针对低延迟应用需求设计了一种Web前端与边缘计算协同的三层架构，该架构包括终端Web前端层，缘计算层与云服务层，核心理念是就近处理分层协作^[4]。架构涵盖前端智能分流、边缘就近处理和云端统筹管理三大核心组件。前端智能分流是指Web前端依据任务特性、网络状况及用户交互模式，对计算任务的执行位置进行智能决策；边缘就近处理则将延迟敏感型计算任务分配至临近的边缘节点处理，其中包含低延迟导向的服务发现与选择机制、基于WebAssembly的边缘执行环境、实时构建的边缘资源地图，以及基于地理位置的粗粒度预筛选；云端统筹管理负责复杂任务与全局优化工作，涉及全局资源调度与优化、复杂计算任务处理、系统状态统筹管理，以及跨区域协调。

三层之间通过统一的服务发现，任务调度与状态同步机制协同工作，形成完整的低延迟处理闭环，为实现高效协同，架构采用了基于加权评分的执行位置决策模型如下。

$$\text{Score}(t,n) = w1 * \text{Perf}(t,n) + w2 * \text{Latency}(t,n) + w3 * \text{Cost}(t,n)$$

其中t表示任务，n表示节点（本地，边缘或云端），Perf、Latency与Cost分别表示性能评分、延迟评分与资源消耗评分，w1、w2与w3为权重系数。

二、关键技术实现与优化

（一）前端感知的计算任务划分

前端计算任务划分是实现低延迟架构的基础，针对Web应用特点，设计了一套计算图分析与任务划分技术，包括依赖分析、粒度控制以及状态管理三个主要方面，计算任务划分采用了改进的HEFT（Heterogeneous Earliest Finish Time）算法，考虑了Web前端特有的DOM操作与渲染因素^[5]。开发了自动化的应用分析工具，集成于前端构建流程，可自动生成任务划分配置，以图像处理应用为例，系统将完整的处理流程分解为图像获取预处理，特征提取模型推理与结果渲染等任务单元，并根据计算复杂度与数据依赖关系，确定每个任务的最优执行位置。

（二）边缘计算节点服务发现与选择

为实现Web前端与边缘节点的高效协作，设计了低延迟导向的服务发现与选择机制，该机制结合DNS-SD，mDNS与中心化目录服务，实现边缘节点的多维度发现，实时采集边缘节点的延

迟负载以及可用资源信息，综合考虑网络距离，负载与服务质量选择最优边缘节点。

系统采用WebRTC STUN服务进行网络探测，结合边缘节点健康度API，构建了实时的边缘资源地图，系统还支持基于地理位置的粗粒度预筛选，降低服务发现的开销。

（三）WebAssembly边缘执行环境

为解决边缘节点上执行Web前端卸载任务的效率问题，设计了基于WebAssembly的边缘执行环境，该环境基于Wasmtime实现轻量级运行时，支持WebAssembly模块的高效执行，通过WebAssembly System Interface提供文件系统，网络等系统级能力，实现基于命名空间的多租户隔离保证不同应用间的安全性。

该执行环境通过以下优化提升性能，AOT编译预先将WebAssembly编译为本地代码，减少启动时间，内存池管理实现高效的内存分配与回收机制，SIMD加速利用WebAssembly SIMD扩展加速向量计算，测试显示相比传统的JavaScript执行环境，WebAssembly边缘执行环境在计算密集型任务上性能提升了3.2倍，内存占用减少了40%。

（四）前端与边缘节点实时通信

为保证前端与边缘节点间的低延迟通信设计了多通道实时通信机制，包括基于QUIC协议的WebTransport通道，提供可靠双向通信的WebSocket通道与用于单向通知的SSE通道，针对不同类型的业务数据，系统自动选择最适合的通道，控制指令采用WebSocket确保可靠性，大块数据采用WebTransport提高传输效率，状态更新采用SSE降低协议开销^[6]。同时系统设计了针对移动网络特性的适应性通信方式，实时监测网络状况并调整通信参数，在网络波动时保持会话状态实现平滑恢复，并根据数据重要性网络状况调整压缩以及发送方式，通过多通道通信机制，系统在不同网络条件下都能保持较低的通信延迟。

三、实验评估与应用场景

（一）实验环境与性能分析

为验证协同架构的有效性，构建了包含多种网络环境与应用场景的测试平台，测试环境包括不同规格的终端设备，分布在多个地理位置的边缘节点与区域中心的云服务器，网络环境覆盖WiFi4G以及5G等多种条件，测试应用包括图像识别实时视频分析与交互式3D渲染三类典型低延迟应用。

延迟性能测试结果显示，在图像识别应用中协同架构比传统云计算架构降低了58%的延迟，比纯边缘计算架构降低了21%的延迟，在不同网络条件下协同架构表现出更好的适应性。特别是在网络质量波动时，延迟抖动明显小于对比架构，在移动设备上协同架构的延迟优势更为显著，平均降低了63%的处理时间。

用户体验测试采用客观指标评估与主观打分两种方法，在实时视频分析应用中协同架构的MOS评分为4.3（满分5分），比传统云架构高0.8分，在交互式3D渲染应用中93%的用户认为协同架构提供了流畅且无明显延迟的体验，通过A/B测试比较不同架构，用户对协同架构的满意度提升了43%，特别是在移动网络

环境下满意度提升更为明显。

（二）典型应用场景分析

1. 实时视频分析应用

实时视频分析是边缘计算的典型应用场景，在协同架构下视频分析应用的工作流程包括：Web 前端通过摄像头 API 采集视频并进行基本预处理，将视频帧传输至边缘节点进行目标检测与跟踪等计算密集型处理，边缘节点将分析结果返回前端进行可视化渲染与用户交互处理^[7]。协同架构在实时视频分析场景下，相比传统云架构可将端到端延迟从450ms降低到180ms，达到实时交互的要求，应用实施中的关键优化包括视频帧差分传输，模型分层推理以及预测性缓存。

2. 交互式边缘游戏

交互式边缘游戏是对低延迟架构的严峻挑战，协同架构在游戏场景中的实现包括：前端捕获用户输入并进行本地状态预测，边缘节点执行游戏主逻辑，云端维护全局游戏状态^[8]。测试结果显示，协同架构使游戏的输入响应延迟从传统架构的120ms降低到45ms，大幅提升了游戏体验，实现中的技术要点包括输入预测与回滚机制，区域边缘节点协同与 WebGPU 渲染优化。

四、成果与创新点

（一）主要成果

Web 前端与边缘计算协同的低延迟应用架构取得了以下几项主要成果：通过前端智能分流边缘、就近处理、云端统筹管理三大核心机制，实现了计算任务的最优分配，架构评估显示相比传统云计算架构可降低58%的端到端延迟，显著提升了实时交互应用的性能^[9]。关键技术突破在 WebAssembly 边缘执行环境，前端感知的计算任务划分，边缘计算节点服务发现与选择以及多

通道实时通信机制等关键技术取得突破，为协同架构的实现提供了技术支撑，特别是 WebAssembly 边缘执行环境，相比传统 JavaScript 环境在计算密集型任务上性能提升了3.2倍。

（二）创新点总结

主要创新点包括前端与边缘协同机制，首次提出并实现了 Web 前端与边缘计算深度协同的机制，打破了传统前后端分离的开发模式，将前端视为分布式计算的重要参与者，而非简单的展示层，层次计算分流方式，创新性地设计了基于任务特征，网络状况以及设备能力的多层次计算分流方式，实现了计算任务在前端边缘与云端的动态优化分配^[10]。WebAssembly 边缘执行环境，开发了适用于边缘计算场景的 WebAssembly 执行环境，解决了边缘节点上高效执行 Web 前端卸载任务的技术难题，为协同架构提供了统一的计算基础，实时通信优化技术，提出了针对前端与边缘节点间的多通道实时通信机制，结合 WebTransport、Web-Socket 与 SSE 等技术，实现了不同类型业务数据的优化传输，降低了通信延迟。

五、结语

通过 Web 前端智能分流技术、边缘节点就近处理机制与云-边-端资源协同调度算法三大核心技术，成功构建了跨越前端、边缘以及云端的分层协同架构，有效解决了地理分布式应用的延迟瓶颈问题，基于 WebAssembly 的边缘执行环境，前端感知的计算任务划分与多通道实时通信机制等关键技术的突破，为协同架构提供了坚实的技术基础，未来随着5G/6G网络的普及，物联网设备的增加与实时交互需求的增长，Web 前端与边缘计算协同架构将发挥更加重要的作用。

参考文献

- [1] 胡强, 张娜, 朱珍元. Web 端与移动端应用平台技术融合的研究 [J]. 芜湖职业技术学院学报, 2023, 25(04): 54-58.
- [2] 张羽佳, 赵爱美. “1+X”证书制度背景下软件技术专业人才培养模式改革与探索——以 Web 前端开发职业技能等级证书为例 [J]. 营销界, 2023, (23): 125-127.
- [3] 基于5E模式下《web 前端》课程思政建设的研究与实践 [C]// 百色学院马克思主义学院, 河南省德风文化艺术中心. 2023年高等教育科研论坛桂林分论坛论文集. 陕西邮电职业技术学院, 2023: 86-87.
- [4] 宁玉丹. 基于“岗课赛证”融通的“Web 前端开发”课程教学改革的研究 [J]. 信息系统工程, 2023, (11): 153-156.
- [5] 张毅宇, 徐梦雨, 马建勇. 软件工程中 Web 开发技术的应用研究 [J]. 中国高科技, 2023, (20): 120-122.
- [6] 周沐玲. 高并发访问下的移动 Web 前端浏览性能优化研究 [J]. 重庆科技学院学报 (自然科学版), 2023, 25(05): 63-68.
- [7] 周香庆, 龙江腾. “工单制”教学模式在 Web 前端开发课程中的应用研究 [J]. 电脑知识与技术, 2023, 19(28): 178-180.
- [8] 陈红卫. 基于 HTML5+CSS3 的 Web 前端响应式页面布局研究 [J]. 数字通信世界, 2023, (09): 26-28.
- [9] 陈美伊. Web 前端开发基础课程思政标准化教学路径研究 [J]. 中国标准化, 2023, (18): 213-215.
- [10] 葛进, 陆雪松. 基于 UI 图像的 Web 前端代码自动生成 [J]. 华东师范大学学报 (自然科学版), 2023, (05): 100-109.