

云计算环境下网络延迟对应用服务质量的影响及对策

叶春艳

杭州萧山技师学院, 浙江 杭州 310014

DOI: 10.61369/ERA.12283

摘 要 : 本文聚焦云计算环境下的网络延迟问题, 详细阐述了网络延迟的定义、测量方法及来源。深入分析其对应用服务质量的影响, 涵盖应用服务质量的评价指标以及对不同类型应用的具体影响。提出了减少网络延迟的技术手段与提高应用服务质量的策略。通过某知名在线教育平台案例, 展现网络延迟引发的问题及应对策略实施后的显著效果, 为云计算服务提供商和用户提供应对网络延迟问题的参考。

关 键 词 : 云计算; 网络延迟; 应用服务质量; 边缘计算; 案例研究

Impact of Network Latency on Application Service Quality in Cloud Computing Environment and Countermeasures

Ye Chunyan

Hangzhou Xiaoshan Technician College, Hangzhou, Zhejiang 310014

Abstract: This article focuses on the issue of network latency in a cloud computing environment, elaborating on the definition, measurement methods, and sources of network latency. It provides an in-depth analysis of its impact on application service quality, covering evaluation metrics of application service quality and specific impacts on different types of applications. Technical means to reduce network latency and strategies to improve application service quality are proposed. Through a case study of a renowned online education platform, the article demonstrates the problems caused by network latency and the significant effects after implementing countermeasures, providing a reference for cloud computing service providers and users to address network latency issues.

Keywords: cloud computing; network latency; application service quality; edge computing; case study

引言

在数字化与信息化迅猛发展的当下, 云计算是企业 and 机构处理海量数据、支撑复杂应用的关键。但其复杂和分布式特性带来诸多挑战, 网络延迟严重影响应用服务质量。网络延迟是数据从发送端到接收端的耗时, 直接左右应用响应速度与用户体验, 尤其在在线游戏、视频会议、金融交易等实时数据处理应用中问题突出。网络延迟不仅关乎用户体验, 对企业运营和竞争力影响巨大, 像金融行业实时数据交易对延迟要求极高, 微秒级差异都可能造成巨额损失。

一、云计算环境下的网络延迟问题

(一) 网络延迟的定义与测量

网络延迟是指数据包从发送端传输到接收端所需的时间, 通常包括传输延迟、处理延迟、排队延迟和传播延迟等几个部分。传输延迟是指数据包在实际传输媒介中传播所需的时间, 这与传输距离和介质特性有关。处理延迟是数据包在网络设备(如路由器和交换机)中进行处理所需的时间。排队延迟是数据包在网络设备中等待处理所需的时间, 这通常与网络拥塞有关^[1]。传播延迟是数据包在网络中传播所需的时间, 这与网络的拓扑结构和数据传输速率有关。测量网络延迟通常采用多种方法, 如 ping 命令、traceroute 工具和网络性能监测软件等。Ping 命令通过发送

ICMP 回显请求包并记录响应时间来测量延迟。Traceroute 工具通过记录数据包到达目标所经过的每个网络节点的延迟, 帮助识别延迟较高的网络段。网络性能监测软件则可以实时监测网络延迟, 并提供详细的性能报告。

(二) 网络延迟的来源分析

在云计算环境中, 网络延迟来源多样。物理距离是基础因素, 数据传输距离越远, 延迟往往越高。网络设备性能至关重要, 低性能的路由器、交换机等, 易造成较高的处理延迟与排队延迟。网络拥塞也是常见成因, 当网络数据流量超处理能力, 数据包需排队, 延迟随之增加。数据中心内部架构也影响网络延迟, 复杂架构使数据包在内部传输时, 要经过多个网络设备, 从而增加传输延迟。数据传输协议的选择同样关键, 不同协议处理

效率与传输延迟有别。像 TCP 协议靠建立连接和确认机制保障数据可靠传输，却易导致较高延迟；UDP 协议不设连接建立和确认机制，传输延迟低，可靠性欠佳。此外，云计算环境也会对网络延迟产生影响。首先，云计算资源主要部署在远程的数据中心，将巨量的分布式数据传输到远程数据中心将使得数据应用经历较长的网络延迟。特别地，随着移动应用的不断增长，不断增长的网络负载将导致不可容忍的网络延迟。其次，从终端用户将巨量的数据传输到远程云计算中心的方式也消耗了巨量的网络带宽资源和传输能量^[2]。

二、网络延迟对应用服务质量的影响

（一）应用服务质量的评价指标

评价应用服务质量的关键指标涵盖多个重要维度，其中包括响应时间、吞吐量、可用性以及用户满意度。响应时间从用户发起请求的瞬间开始计时，直至用户成功接收响应，这段时长直观反映应用性能。像在线购物平台，更短响应时间让消费者能快速获取商品信息、完成交易，极大提升购物体验。吞吐量代表单位时间内应用处理的最大请求数量，彰显应用承载能力，电商大促时，高吞吐量应用系统可同时应对海量用户访问与下单，保障平台平稳运行，避免卡顿崩溃。可用性是应用在特定时间段正常运行的占比，对金融交易类应用极为关键，交易时段稳定运行才能保障资金安全与交易顺畅。用户满意度从用户视角主观评价应用服务质量，通过收集用户使用反馈或开展调查获取，有助于全方位了解用户体验，进而持续优化服务质量。

（二）网络延迟对不同应用的影响分析

网络延迟对不同类型的应用影响程度不同。对于实时性要求高的应用，如在线游戏和视频会议，网络延迟会直接影响用户体验。例如，在线游戏中，玩家操作的实时反馈对游戏体验至关重要，较高的网络延迟会导致操作滞后，严重影响游戏的可玩性。视频会议中，延迟会导致通话双方交流不畅，影响沟通效果。在金融交易系统中，网络延迟可能导致交易指令的执行滞后，影响交易效率和准确性。在电子商务平台中，网络延迟会影响用户的购物体验，较高的延迟可能导致用户放弃交易，影响平台的销售额^[3]。对于数据分析和处理应用，虽然网络延迟的影响可能不如实时应用那么直接，但它仍然会影响数据处理的速度和效率。例如，在大数据分析中，延迟较高的网络可能会导致数据处理任务的完成时间延长，影响分析结果的及时性和准确性。

三、应对网络延迟的策略与优化措施

（一）优化网络架构

精简数据传输路径，减少不必要的网络设备跳转。例如，通过合理规划网络拓扑，使数据在从用户端到云服务器的传输过程中，尽量减少经过路由器、交换机等设备的数量。这样做能有效降低数据包在网络设备中的处理延迟和排队延迟，加快数据传输速度。同时，采用高速网络链路，如升级到万兆以太网甚至更高

速的光纤网络，提高数据传输的物理速率，缩短传输延迟^[4]。并且定期对网络设备进行性能评估与更新换代，淘汰老旧、性能低下的设备，保证网络设备能高效处理数据流量，进一步减少延迟产生的可能性。

（二）利用内容分发网络（CDN）

CDN 将内容缓存到离用户更近的边缘服务器。当用户请求数据时，优先从距离最近的 CDN 节点获取，而非直接从远程数据中心读取。以视频网站为例，热门视频内容可提前分发至分布在各地的 CDN 服务器上。用户观看视频时，数据从本地附近的 CDN 节点传输，大大缩短了传输距离，降低了网络延迟，使视频播放更加流畅，减少卡顿现象。CDN 还具备负载均衡功能，能将大量用户请求分散到多个节点处理，避免单个服务器因负载过高而产生延迟。

（三）采用高效数据传输协议

传统的 TCP 协议在确保数据可靠传输方面表现出色，但因其建立连接和确认机制复杂，往往会带来较高延迟^[5]。可引入更高效的传输协议，如 QUIC（快速 UDP 互联网连接）协议。QUIC 基于 UDP 协议，减少了连接建立的往返次数，能更快地建立数据传输通道。同时，它在传输过程中对数据包的处理更加高效，可有效减少数据传输时间，提升整体传输效率，降低网络延迟。尤其在实时性要求高的应用场景，如在线游戏、视频会议中，QUIC 协议能显著改善用户体验。

（四）借助边缘计算技术

将部分计算资源部署到靠近数据源或用户端的网络边缘。比如在物联网应用中，大量传感器产生的数据无需全部传输到远程云数据中心处理，可在边缘节点直接进行初步处理。这样，一方面减少了数据在网络中的传输量，降低了传输延迟；另一方面，边缘节点能快速响应用户请求，提高应用的响应速度。以智能工厂为例，车间内的设备数据在边缘计算设备上实时分析与处理，及时反馈设备运行状态，快速做出控制决策，避免因数据传输到远程云中心产生延迟而影响生产效率^[6]。

（五）基于软件定义网络（SDN）优化

SDN 通过将网络的控制平面与数据平面分离，实现网络的集中控制与灵活配置。网络管理员可根据实时网络流量情况，在控制平面通过软件编程的方式，动态调整数据平面的流量转发路径。当某条链路出现拥塞时，SDN 可迅速将流量切换到其他空闲或负载较低的链路，减少网络拥塞导致的延迟。同时，SDN 能够对网络资源进行精细化管理，合理分配带宽等资源，确保关键应用和实时性要求高的业务获得足够的网络资源，保障其低延迟运行。

（六）增强应用设计与优化

在应用开发阶段，优化代码结构，减少冗余代码，提高代码执行效率。例如，采用高效的算法和数据结构处理数据，避免复杂的循环和递归操作，以降低应用处理数据的时间。对频繁访问的数据进行缓存，可在本地内存或分布式缓存系统中存储常用数据，当用户再次请求相同数据时，直接从缓存中读取，减少从数据库或远程服务器获取数据的时间，从而降低应用的响应延迟^[7]。

并且对应用进行分层架构设计，各层职责明确，相互独立，便于进行针对性的优化和扩展，提高应用整体的性能与稳定性，减少因应用自身问题导致的延迟。

（七）实时监控与用户反馈机制

部署专业的网络性能监测工具，实时监测网络延迟、带宽使用情况、服务器负载等关键指标。通过对这些数据的实时分析，及时发现网络中存在的延迟问题及潜在风险。例如，当监测到网络延迟突然升高时，可迅速定位是网络链路故障、服务器负载过高还是其他原因导致，并及时采取相应措施，如修复链路故障、增加服务器资源等。同时，建立有效的用户反馈机制，鼓励用户报告在使用应用过程中遇到的延迟问题^[9]。通过用户反馈，能更直观地了解应用在实际使用场景中的性能表现，结合监测数据进行综合分析，进一步优化应用和网络服务，提升用户体验。

四、案例研究

（一）案例背景

某家知名在线教育平台，为全球数百万学生提供各类课程直播、在线辅导以及学习资料下载等服务。平台采用云计算架构，将课程资源存储在位于美国的数据中心，以服务全球用户。随着业务的快速扩张，尤其是在亚洲地区用户数量的激增，平台开始频繁收到用户反馈，称在上课过程中出现视频卡顿、资料下载缓慢以及与教师互动延迟等问题，严重影响了学习体验，导致部分用户流失，平台口碑也受到一定影响。

（二）网络延迟问题分析

在平台服务中，亚洲地区用户因与美国数据中心距离远、传输链路长，面临严重网络延迟。这对课程直播、资料下载及互动环节都产生了负面影响。课程直播时，视频数据传输受阻。例如面向日本学生的语言直播课，平均每10分钟就会卡顿3—4次，每次卡顿5~10秒，打乱教学节奏，学生难以跟上教师授课，学习效果不佳^[9]。课后资料下载也深受其害。正常1~2分钟就能完成的100MB资料下载，受网络延迟影响，东南亚地区用户下载时间延长至10~15分钟，耗费大量时间，可能使学生放弃下载，打乱学习计划。在线教育关键的互动环节同样未能幸免。在印度学生参与的在线辅导课上，学生提问后教师平均要等8~10秒才能看到，教师回复后学生也需等类似时长才能收到，这种延迟严重阻碍师生有效沟通，极大降低课堂互动的积极性与效率，对在线教育的教学质量造成严重冲击。

（三）应对策略与实施效果

为解决网络延迟对在线教育服务质量的影响，该平台采取多项有效举措。在亚洲的新加坡、日本、印度等地部署边缘计算节点，缓存热门课程与常用资料，学生能从更近节点获取数据，日本地区课程视频卡顿减少80%，资料下载速度提升5倍，平均下载时间缩至2~3分钟。平台运用并行计算技术，将视频转码等复杂任务分割，在多台云服务器并行处理，直播课程视频处理速度提高3倍，直播更流畅。采用RCLLOUD算法，借助DHT技术和随机邻居策略，印度地区学生搜索资料时间平均缩短60%^[10]。依据

不同地区的网络状况和用户习惯分组，为网络好的韩国用户组采用高效传输协议、优先分配带宽，为网络不稳定的东南亚部分地区用户组优化数据缓存策略，实施后整体用户服务质量满意度提升30%，网络延迟问题得以有效缓解。

五、结束语

本文深入剖析云计算环境下网络延迟及其对应用服务质量的影响，探讨了相关优化策略并以案例验证效果，为云计算服务提供商和用户提供参考。展望未来，随着云计算技术发展与新应用出现，网络延迟仍是研究重点。5G的高速低延迟、物联网设备产生的海量数据，以及边缘计算在减少传输延迟上的优势，都对云计算网络延迟管理提出新的挑战与机遇。未来研究可聚焦于融合5G、物联网、边缘计算等新技术，进一步优化网络延迟、提升服务质量，同时探索运用人工智能和机器学习技术，实现网络延迟的智能化管理，全面提升云计算服务性能。

参考文献

- [1] 禹晶晶. 基于延迟传输模式的无线体域网同传技术研究[D]. 吉林: 吉林大学, 2020.
- [2] 郭棉, 李琦琦. 面向延迟敏感型物联网应用的计算迁移策略[J]. 计算机应用, 2019, 39(12): 3590-3596.
- [3] 崔建明, 刘佳伟, 杨呈永. 基于优先排队论网络延迟云计算资源调度算法[J]. 桂林理工大学学报, 2017, 37(2): 360-365. DOI: 10.3969/j.issn.1674-9057.2017.02.019.
- [4] 许鸿柳. 基于移动边缘云计算中部分计算迁移的延迟感知资源分配[D]. 四川: 电子科技大学, 2023.
- [5] 梁成龙. 面向低延迟网络数据传输的FPGA系统设计[D]. 江苏: 南京邮电大学, 2023.
- [6] 邹劲松, 唐旭. 适用于自适应网络防御的联合云计算框架[J]. 西南师范大学学报(自然科学版), 2020, 45(3): 121-126. DOI: 10.13718/j.cnki.xsxb.2020.03.019.
- [7] 纪梓潼. 云计算环境混合负载下尾延迟性能的评测与分析[D]. 天津: 天津大学, 2018.
- [8] SEYEDEH MAEDEH MIRMOHSENI AMIRI. 基于优化技术的云计算网络资源配置改进[D]. 广东: 广州大学, 2021.
- [9] 苏州大学. 边缘云计算的弹性光网络最大容忍延迟再分配方法与系统: CN202211346546.3[P]. 2024-11-01.
- [10] 严立宇, 祖立军, 叶家伟, 等. 云计算网络中多租户虚拟网络隔离的分布式实现研究[J]. 计算机应用与软件, 2016, 33(11): 93-98. DOI: 10.3969/j.issn.1000-386x.2016.11.022.