**文章编号**：1003-1421(2026)03-0177-11
**引用格式**：张智，樊春美，李雯，等.铁路互联网售票系统智能流量调度研究[J].铁道运输与经济，2026，48(3)：177-187.
ZHANG Zhi，FAN Chunmei，LI Wen，et al. Intelligent Traffic Scheduling for Railway Online Ticketing Systems[J]. Railway Transport and Economy，2026，48(3)：177-187.

# 铁路互联网售票系统智能流量调度研究

张　智，樊春美，李　雯，宋宇光，仲　硕

(中国铁道科学研究院集团有限公司 电子计算技术研究所，北京　100081)

**摘　要**：随着5G应用推动互联网售票流量激增，传统静态规则的调度算法难以应对多数据中心在动态网络、异构资源与服务多样性方面的挑战。因此，设计集内容分发网络和统一接入平台的多层智能流量调度系统，提出了基于机器学习算法的多数据中心流量调度方法和数据中心内部的流量调度算法，在跨数据中心，采用基于神经网络的预测模型分析网络状态与业务负载，并结合网格搜索方法动态优化流量调度比例，实现多中心间流量的全局均衡与效率最优。在数据中心内部，提出基于基数树的高效路由算法，通过对服务标识进行前缀树构建与匹配，实现海量并发请求的低延迟、高精度转发。该智能流量调度系统及算法，能够提高流量调度效率和精准性，有效保障系统的稳定运行和应急处置。

**关键词**：铁路互联网售票系统；智能流量调度系统；高并发处理；负载均衡；机器学习
**中图分类号**：U293.1　　**文献标识码**：A　　**DOI**：10.16668/j.cnki.issn.1003-1421.20250513002

# Intelligent Traffic Scheduling for Railway Online Ticketing Systems

ZHANG Zhi，FAN Chunmei，LI Wen，SONG Yuguang，ZHONG Shuo

(Institute of Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081,China)

**Abstract**: As the applications of 5G technology lead to a significant surge in online ticketing traffic, traditional scheduling algorithms based on static rules struggle to address the challenges facing multiple data centers in terms of dynamic networks, heterogeneous resources, and service diversity. Therefore, this paper designed a multi-tier intelligent traffic scheduling system integrating the content delivery network (CDN) and unified access platform, and proposed a multi-data-center traffic scheduling method based on machine learning algorithms and an intra-data-center traffic scheduling algorithm. In inter-data-center scenarios, a neural network-based prediction model was adopted to analyze network status and traffic load. Meanwhile, by combining the grid search method, the traffic scheduling ratios were dynamically optimized to achieve global traffic balance and optimal efficiency across multiple data centers. For intra-data-center scenarios, an efficient routing algorithm based on the radix tree was proposed. By constructing and matching prefix trees for service identifiers, the low-latency and high-precision forwarding of massive concurrent requests was achieved. The proposed intelligent traffic scheduling system and algorithm can improve both traffic scheduling efficiency and precision, and ensure the stable operation and emergency response abilities of the system.

**收稿日期**：2025-05-13　　**修订日期**：2025-12-19
**通信作者**：仲硕(1991—)，女，黑龙江哈尔滨人，中国铁道科学研究院集团有限公司电子计算技术研究所工程师。
**基金项目**：中国国家铁路集团有限公司科技研究开发计划课题(P2024S002)

**Keywords:** Railway Online Ticketing System; Intelligent Traffic Scheduling System; High-Concurrency Processing; Load Balancing; Machine Learning

## 0　引言

近年来，随着我国高速铁路网络的迅猛扩展与互联网技术的深度普及，铁路互联网售票系统已成为旅客购票的主要渠道。尤其在高峰期，如春运、节假日等时段，12306平台的单日访问量高达2 600亿次，系统面临极高的流量压力。然而，传统流量调度手段在应对如此大规模并发访问时显得力不从心，难以确保系统始终处于最佳运行状态，进而影响了系统的响应速度与用户体验。因此，探索智能流量调度系统技术，以提升铁路互联网售票系统在高并发环境下的处理能力与资源利用效率，不仅具有重要的理论价值，更有迫切的现实意义。

当前，铁路互联网售票系统正面临一系列严峻挑战：一是超高并发访问已成常态，春运、节假日等特殊时期流量激增，远超日常负荷；二是恶意请求的频繁侵扰，黄牛党利用自动化工具实施大规模刷票，严重挤占了正常用户的资源空间；三是资源分配的不均衡性，静态负载均衡策略难以适应动态变化的流量模式，导致部分服务器长期过载；四是用户体验优化迫在眉睫，如何在高负载状态下保障支付、查询等关键业务的稳定性，成为亟待解决的关键问题。因此需要建立一个面对流量压力大、系统故障时能灵活调整流量的框架，提供多元化流量调度方式的技术支持，以保障系统处于最优的状态，保持系统的稳定性。

在流量调度领域，研究聚焦于策略研究、传统算法如加权轮询、加权公平队列、最短作业优先等方面。刘俊等[1]对铁路运输调度现代化发展策略进行了系统性研究，为宏观调度体系优化提供了理论支撑；张克云等[2]面向发达高速铁路网，提出了调度管理的优化策略，为大规模网络下的资源协同调度提供了重要参考；金子龙等[3]通过预测未来流量并动态调整调度权值，有效降低了时延与丢包率；郑凌等[4]提出了基于时间感知的最大匹配流量调度算法，实现了逐帧最大权重调度匹配；高明明等[5]对朔黄铁路调度行车指挥智算平台的总体架构进行了研究，为铁路调度系统的智能化升级提供了实践路径；雷其鸣[6]提出了基于堆叠学习驱动的种群初始化机制，并结合自适应权重调整与遗传操作，形成了改进强化学习的求解算法；陆以勤[7]结合路由与流量调度约束，设计了路由优化遗传算法，提高了调度算法的求解效率；黄宗伟等[8]针对数据中心网络中的性能瓶颈，聚焦于“大象流”的调度问题，研究了基于SDN的负载均衡策略以提升网络效率；陈杰等[9]从满足多样化网络需求出发，设计了优先级队列流量调度算法，旨在通过对不同业务流的差异化处理来保障关键业务的服务质量；崔羽飞等[10]重点探讨了异构计算资源的统一调度与跨天地的流量卸载技术。

既有研究发现通过深度融合大数据分析、机器学习等前沿技术，不仅能够实现动态流量预测、实时资源分配与恶意请求识别，还能结合用户行为分析，优先保障正常用户的访问权益，提升系统整体公平性，智能流量调度系统技术已在网络架构优化[11-12]、时间敏感网络[13-15]、基于感知与学习的动态调整[16-17]等多个领域展现出广泛应用前景。因此，研究聚焦铁路互联网售票系统的智能流量调度系统问题，结合用户访问层和系统内部流量调度的需求，提出一种智能高效、稳定可靠的流量调度框架，为铁路互联网售票系统的稳定运行提供有力支持。

## 1　智能流量调度系统设计

12306平台的流量高峰具有显著的周期性和突发性特征，在春运、暑运、十一等重大节假日期间以及每日各个放票时段，存在着高峰期业务需求，这些业务主要包括高并发抢票、实时余票查询、突发性退改签、人脸识别等场景，对系统的响应速度、稳定性和数据一致性提出了极高要求。此外，列车晚点或极端天气等突发事件也会导致退改签请求爆发式增长。为应对这些挑战，研究设计面向客运业务的智能流量调度系统。该系统重点保障12306核心售票业务的稳定运行，通过内容分发网络层(Content Delivery Network，CDN)流量调度应对突发购票流量，利用跨数据中心调度实现灾备容灾，借助集群级调度优化资源利用率。

在业务层面，该系统实现了以下三大核心功能。①大客流应对：通过双CDN智能切换，有效缓解春运等高峰时段的区域性访问压力，确保购票流程顺畅；②业务连续性保障：基于跨数据中心调度，在单数据中心故障时实现秒级流量切换，避免售票服务中断；③资源优化：通过集群级动态负载均衡，根据实时业务需求(如余票查询、订单提交等)自动分配算力，提升资源使用效率。

系统通过统一的运维平台实现全链路监控，快速定位并处置网络抖动、服务器过载等影响客运服务的故障，平均故障恢复时间缩短至分钟级，显著提升了旅客购票体验和系统可靠性。

针对铁路互联网售票系统的CDN流量调度、跨数据中心流量调度、数据中心内集群流量调度，设计了铁路互联网售票系统流量调度框架如图1所示，整体架构分为2层，第1层是CDN层，负责实现双CDN的流量切换以及单CDN对不同数据中心的流量调度；第2层是统一接入平台，主要完成同数据中心不同应用以及跨中心应用的流量调度，并通过研发运维平台进行统一的流量监控、故障发现、应急处置、故障恢复等流量调度管理。

### 1.1　内容分发网络层

系统设计了双CDN架构，实现CDN层的应用双活，避免单点CDN服务故障影响整体服务，全面提升业务容灾能力和用户购票体验。

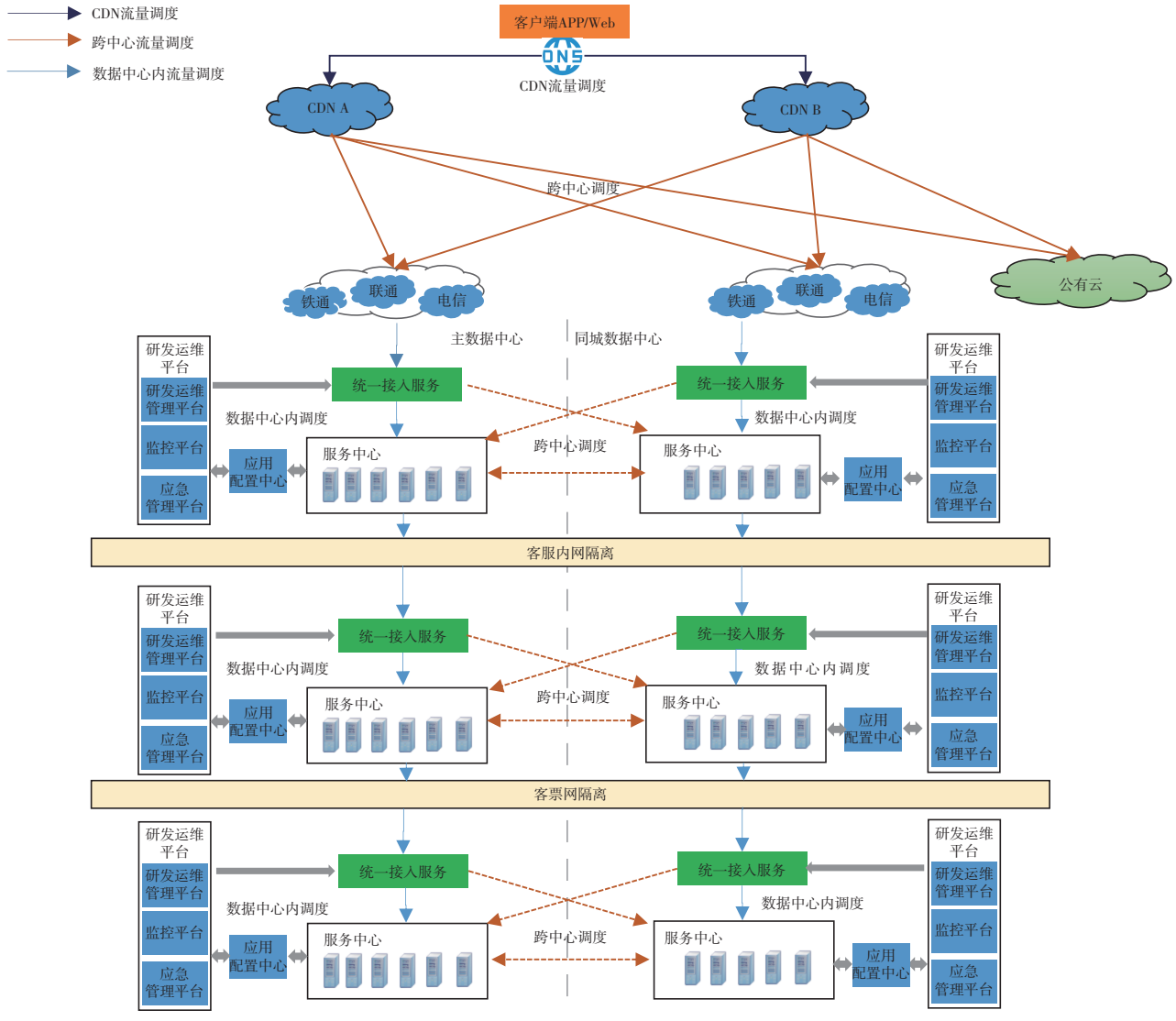


**图1　铁路互联网售票系统流量调度框架**

Fig.1　Traffic scheduling framework for railway online ticketing system

### 1.1.1　CDN双活流量调度

（1）基于DNS调度的双活方案。利用DNS的调度能力对CDN进行流量牵引，按照区域运营商进行回源流量调度策略配置设计，通过DNS实现双CDN中心流量的分配，区域运营商回源流量调度过程如图2所示。

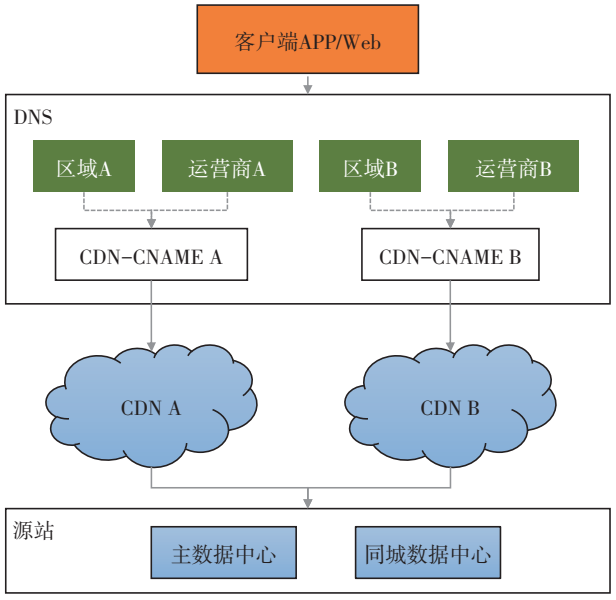


**图2　区域运营商回源流量调度过程**

Fig.2　Origin-bound traffic scheduling process for local carriers

（2）基于HTTP-DNS的双活方案。HTTP-DNS使用HTTP协议进行域名解析，代替现有基于UDP的DNS协议，域名解析请求直接发送到HTTP-DNS服务器，绕过运营商的Local DNS解析，能够避免Local DNS造成的域名劫持问题和调度不精准，解决Local DNS缓存过长导致故障切换生效慢的问题，HTTP-DNS流量调度策略如图3所示。

由DNS、CDN资源调度中心实现流量调度，通过设置流量调度策略实现基于省份、城市的区域流量调度和运营商的流量调度。当某一CDN厂商、某一区域、某一运营商出现故障时，可通过研发运维平台进行流量调度的应急处置，实现故障隔离，保证业务的连续性和稳定性。

### 1.1.2　单CDN流量调度方案

CDN负责互联网售票系统的流量接入与管理，实现不同数据中心的流量分配。通过监控2个中心的资源、服务器性能等指标，预测出不同数据中心最佳的流量权重比例，设置定时更新或者阈值检测

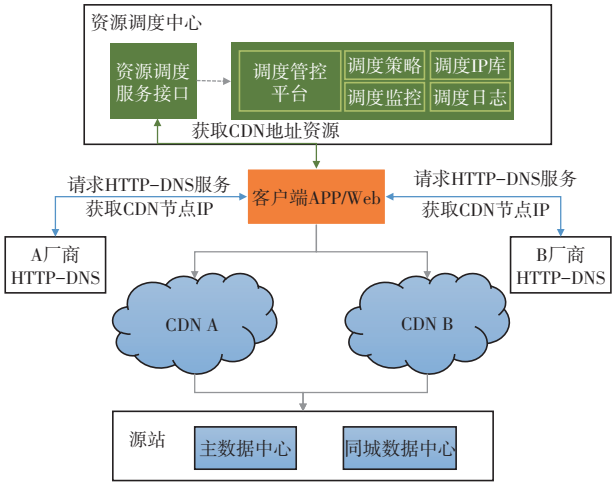


**图3　HTTP-DNS流量调度策略**

Fig.3　HTTP-DNS traffic scheduling strategy

等方式触发流量比例调整，调用CDN更新服务接口API实现各数据中心流量比例权重自动化更新，确保各个数据中心的应用服务性能最优，单CDN流量调度方案如图4所示。以余票查询服务为例，主数据中心与同城数据中心可以按照4:6的权重配比进行流量分配，当同城数据中心服务压力比较大时，要进行流量配比更新，确定一个最优的配比，实现余票服务响应时间最短。

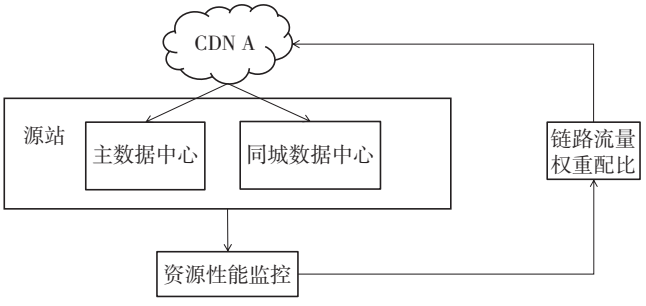


**图4　单CDN流量调度方案**

Fig.4　Single CDN traffic scheduling scheme

## 1.2　统一接入平台

随着售票业务的发展，各类合作平台、路内外系统对接需求显著增加，售票系统建立了多个业务接入的统一接入平台。该平台采用分布式部署与集中管理的架构，为售票系统提供流量接入、精细化管理等能力，支持接入管理、动态路由、服务代理等核心功能，为上层业务提供便捷、可靠的统一接入服务。以进站检票场景为例，其具有高并发、低延迟的特点，在节假日高峰期，大量旅客同时刷身份证/二维码进站，同时检票响应需在毫秒级，否

则可能造成排队拥堵，需要针对流量采取不同的分发策略，甚至可以利用统一接入平台的负载均衡功能进行分级调度。

### 1.2.1　逻辑架构

统一接入平台逻辑架构如图5所示，主要由统一接入边缘节点和统一接入管理门户两部分构成。统一接入边缘节点为流量的代理处理单元，统一接入管理门户为边缘节点的管理单元。统一接入边缘节点通过负载均衡构成统一接入的生产集群，边缘节点响应统一接入管理门户配置指令，支持基于容器运行环境的动态扩容、缩容。统一接入管理门户向边缘节点下发管理控制指令，可跨集群、跨中心调度多个接入集群，提供面向不同用户角色的可视化的管理界面。

### 1.2.2　功能设计

统一接入平台具备HTTP/HTTPS、UDP、TCP等多协议的统一接入能力，提供铁路运输业务标准、统一的认证授权体系，提供运维支持、监控告警、应急管理等能力，针对流量调度应用主要包括接入管理、动态路由、服务代理、运维管理、应急管理等功能。

（1）接入管理。统一接入平台支持多种协议接入，包括HTTP/HTTPS、UDP、TCP等，支持多种接入格式包括文本、JSON、XML以及自定义报文，并可对各种格式和报文进行转换，在余票查询、登录、下单等场景中得到广泛应用。

（2）动态路由。动态路由是统一接入的核心，支持基于网络层协议和应用层协议的路由，支持基于统一资源标识符(Uniform Resource Identifier，URI)的全路径匹配和前缀匹配，能够基于Header、Cookie、Args、UserAgent等作为路由的条件，支持自定义路由匹配函数，支持路由的自动过期(Time To Live，TTL)，支持路由的优先级操作以及路由发布、灰度发布等功能。并且支持异常流量的识别与处理，将不同风险请求转发到相应的服务中，如异常购票请求进入慢速排队系统，不在URL白名单的请求拒绝访问等。

（3）服务代理。服务代理包含服务发现、主动/被动检查、负载均衡等功能。服务发现是通过配置注册中心的地址、服务名称等信息动态获取服务列表信息，实现对动态添加的应用节点进行自动分流；主动/被动检查是通过主动或者被动形式检查后端的应用健康状态，动态剔除故障节点；负载均衡是指统一接入服务在做转发时，对后端服务多节点的分发策略，支持加权轮询、一致性哈希等负载均衡策略。在铁路运输业务中，服务代理技术可实现对列车调度系统、货运追踪服务等关键业务模块的动态治理，负载均衡则优化了余票查询等高峰流量的分发效率，从而支撑铁路业务的高可用与弹性扩展。

（4）运维管理。主要包含日志、监控功能。在日志管理层面，支持可视化的日志配置策略，能够将分布式系统产生的请求日志实时转发至指定的监控系统进行集中分析。同时，系统提供了统一的日志归档功能，可将日志持久化存储于指定文件，并支持日志的在线查看、全文检索与批量下载，极大地提升

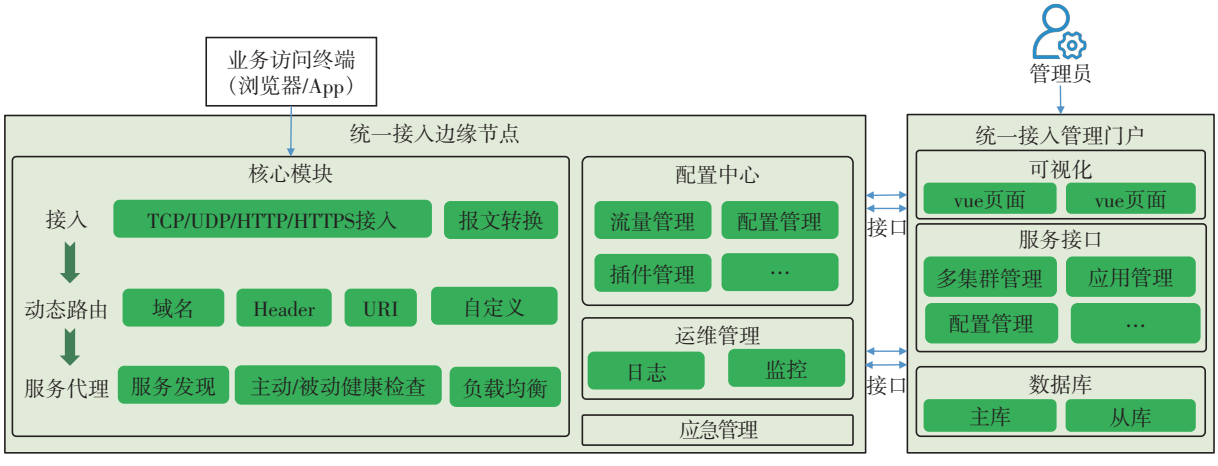


图5　统一接入平台逻辑架构

Fig.5　Logical architecture of unified access platform

了运维排查效率。在监控层面，具备多维度的指标采集与可视化展示能力，能够及时查看流量和HTTP状态码分布。在铁路运输业务中，运维管理的日志与监控功能为运输系统的稳定运行提供关键保障，通过多维监控指标(如客流密度)，实现故障快速定位与资源动态调配，从而确保客运服务的高可用。

（5）应急管理。对接应急管理平台，提供流量管理功能和应急指挥功能。对外提供接口，支持频率控制、流量监控、双中心流量调度和熔断等功能，为应急指挥系统提供多种场景下的应急切换接口，系统本身也支持在线可视化应急操作。

#### 1.2.3　部署架构

统一接入平台系统架构如图6所示，统一接入平台部署在网络的业务边界，3个网络接入平台相互独立，负责本层外部区域流量的接入。在主数据中心和同城数据中心存放异步交易排队数据、营销分析服务、检票服务、实名制等服务，实行数据同步、服务调用、资源共享。三网统一接入平台依据服务系统不同，启用不同功能，客服外网统一接入平台主要对接互联网入口流量，提供高性能的HTTP流量接入和安全防护，侧重于高并发支持、通用安全防护、业务安全保护等；客服内网统一接入平台主要负责从外到内的业务流量处理，提供HTTP、Webservice协议的流量接入；客票网统一接入平台接收客服内网的流量。

铁路互联网售票系统构建了多层次的流量调度框架，通过CDN双活、统一接入平台及智能运维体系，保障高并发场景下的系统稳定性和用户体验。在CDN层采用DNS与HTTP-DNS双调度方案，实现区域及运营商级精准流量分配，有效规避单点故障和域名劫持问题。统一接入平台作为售票系统的核心流量调度中枢，通过多层次、多协议、智能化的功能设计，构建了“接入-路由-代理-监控-应急”的全链路流量调度体系，实现了高效、安全、灵活的流量接入与调度管理。该框架显著提升了系统调度灵活性和服务质量，为春运等峰值流量场景提供可靠支撑。

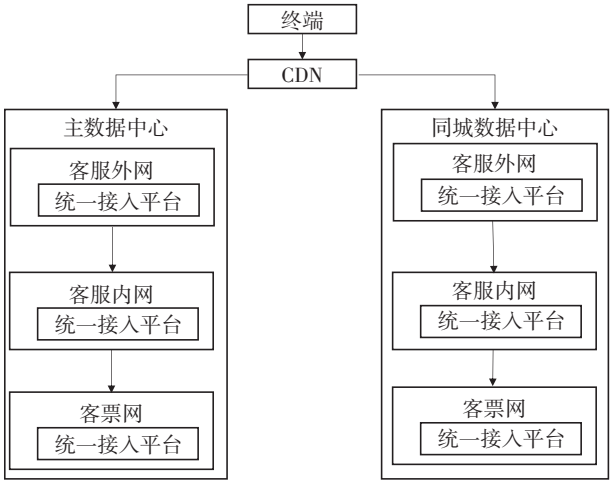


**图6　统一接入平台系统架构**

Fig.6　System architecture of unified access platform

## 2　智能流量调度系统策略

针对铁路售票系统多数据中心下的业务特性，本框架设计了分层协同的智能调度体系，重点解决运输业务中面临的两大核心挑战。①跨中心动态流量调度。基于铁路运输的时空不均衡性问题，如春运期间热门线路的余票查询、订单购票的流量急剧增长等情况，构建了基于神经网络模型的流量调度算法，通过融合系统性能指标动态计算最优流量分配比例。②数据中心内部业务级调度。针对售票业务的服务等级差异(如支付事务优先级高于历史查询)，采用增强型基数树路由算法，依托统一接入平台与应用配置中心的协同管控机制，构建了面向铁路售票业务的精细化流量调度体系，实现基于运输业务特征的动态负载均衡，当检测到某集群异常(如北京西站电子客票服务响应超时)，高可用SDK实时联动应用配置中心，实现故障节点自动移除、备用集群流量接管。以铁路预约购票业务为例，某大区用户流量被DNS引导至主数据中心，当主数据中心CPU超过告警阈值时，会将预约购票请求流量分发至同城数据中心。

### 2.1　跨中心的流量调度

#### 2.1.1　跨中心流量调度

跨数据中心流量调度体系是保障铁路运输服务高可用的核心支撑，通过CDN智能调度全国旅客的12306访问请求至最优数据中心，结合统一接入平台对登录、下单、余票等关键业务的多中心流量动态调配，确保极端情况下(如自然灾害、硬件故障)运输指挥不中断。跨数据中心流量调度由CDN、统一接入平台实现。CDN负责互联网入口流量的接

入管理，通过调整CDN的回源策略，按照运营商、业务规则、系统资源使用率等情况，实现主数据中心、同城数据中心、灾备数据中心、公有云的动态流量分拨。统一接入平台负责客服外网、客服内网、客票网的流量接入管理，当系统出现资源紧张、故障、重大升级、数据中心迁移等情况时，可按照不同业务、服务能力、资源利用率等策略，进行跨中心、跨应用的流量再分配。研发运维平台负责流量调度的集中管理，通过对接CDN厂商、统一接入平台、应用配置中心，实现跨CDN、跨中心、中心内多集群、多应用流量的动态调度。除了支持基于人工经验的调度，还提供完善的自动化调度平台，如基于系统运行状态和流量情况寻优最佳的数据中心流量分配比例，根据全链路监控故障告警事件和智能算法动态生成调度任务，实现基于故障自愈的智能流量调度系统。当某一数据中心出现故障时，研发运维平台的智能调度模块可基于CDN和统一接入平台实现故障自动隔离，保证业务系统的连续性和稳定性。

### 2.1.2　基于神经网络的流量调度算法

当前铁路互联网售票系统部署在主数据中心和同城数据中心，两个中心的资源不同、服务器性能不同，不能简单平均分配2个中心的流量，目前是根据历史运行情况预估比例，为了实现系统整体性能达到最优，确定最合适的流量分配比例，提出基于神经网络的动态流量调度算法，旨在根据系统状态动态调度请求流量。将该问题建模为一个动态优化问题，提出基于神经网络的调度优化框架。首先训练神经网络模型并基于该模型预测响应时间，基于网络搜索确定最优调度比例。

（1）神经网络模型学习。该算法首先基于现有指标数据训练神经网络模型，输入为同城数据中心、主数据中心的CPU利用率平均值与最大值、带宽总入口流量、内存、磁盘占用率、流量占比等系统指标以及调度比例，输出为系统响应时长。所有输入特征在训练前均进行了Z-score标准化处理，以消除量纲差异，加速模型收敛。模型采用多层感知机(Multilayer Perceptron，MLP)结构，有一个输入层、两个隐藏层、一个输出层，输入层包含14个神经元，隐藏层的神经元数量分别为64和32，输出层有1个神经元。隐藏层使用ReLU函数，输出层使用线性激活函数，损失函数采用均方误差(Mean Squared Error，MSE)，以直接拟合响应时间的连续值，使用Adam优化器，其学习率设置为0.001，训练时的批量大小(Batch Size)设置为32。神经网络通过足够的训练数据和适当的超参数设置，能够逼近系统性能函数，实现收敛，找到全局最优解。

对于单个数据样本的一次前向传播，其时间复杂度$O_{\text{forward}}$计算如下。经计算$O_{\text{predict}}$为2 976。

$$O_{\text{forward}} = O\left(n_{\text{input}} \cdot n_{\text{hidden1}} + n_{\text{hidden1}} \cdot n_{\text{hidden2}} + n_{\text{hidden2}} \cdot n_{\text{output}}\right) \tag{1}$$

式中：$n_{\text{input}}$为输入层神经元数量；$n_{\text{hidden1}}$为第一个隐藏层神经元数量；$n_{\text{hidden2}}$为第2个隐藏层神经元数量；$n_{\text{output}}$为输出层神经元数量。

在预测阶段，对于$n$个待预测的样本，需要执行$m$次前向传播。因此，总预测时间复杂度$O_{\text{predict}}$计算如下。

$$O_{\text{predict}} = n \cdot m \cdot O_{\text{forward}} \tag{2}$$

（2）基于神经网络的流量调度学习。获得训练好的预测模型后，算法通过网格搜索(Grid Search)来寻找最优流量分配比例。搜索策略为在连续的调度比例空间[0.20，0.80]范围内，以0.01为步长，生成61个候选比例(0.20，0.21，…，0.80)。寻优过程为在每个调度周期，算法获取最新的系统指标(CPU、内存等)，并将其与每一个候选比例组合成一个新的输入向量。将该向量输入到已训练好的神经网络模型中，得到对应比例下的预测响应时间。最终，选择预测响应时间最短的那个比例作为当前最优调度比例。将此最优比例应用于后续的流量调度。考虑到网格搜索的计算开销，调度寻优的间隔时间设定为5 min。此间隔在保证系统能及时响应状态变化的同时，避免了过于频繁的搜索对系统本身造成额外的计算负担。模型采用固定周期的全量重训练策略，每周自动执行一次完整训练流程。训练数据采用前期90%的训练数据加上最新48 h的数据，确保模型能够学习到最新的系统特征与性能模式。算法实时监控模型预测误差，当连续6个调度周期内平均预测误差超过阈值时，自动触发模型提前重训练。此机制能有效应对因节假日等特殊场景引起的系统特性变化。

### 2.2　数据中心内部的流量调度

#### 2.2.1　数据中心内部流量调度

数据中心内部流量调度由统一接入平台、应用配置中心实现。统一接入平台负责数据中心内部多应用集群的流量分配，应用系统通过集成高可用SDK，实现基于应用配置中心的多集群、多应用的流量均衡调度。当某一应用或集群出现故障时，如人脸集群故障影响旅客检票进站，研发运维平台的智能调度模块可基于统一接入平台和应用配置中心实现故障自动隔离，保证业务系统的连续性和稳定性。下面将详细介绍针对特定服务流量调度的智能调度算法。

#### 2.2.2　特定服务流量调度算法

在统一接入平台中，需要对特定服务的流量进行高效调度与路由，确保请求能够根据服务标识快速定位到对应的服务实例或后端集群。基数树作为一种高效的前缀树结构，能够快速匹配字符串前缀，适合用于服务路由场景。本算法旨在利用基数树实现服务流量的动态调度与流转，支持高并发、低延迟的服务发现与路由。算法核心思想是将服务标识存储为基数树的路径节点，每个节点包含路径前缀、子节点引用及附加元数据，根据请求中的服务标识前缀，在基数树中快速定位目标服务节点，并根据节点元数据选择具体服务实例，支持服务实例的动态增删、权重调整及服务标识的版本升级，确保基数树的实时性。算法流程如下。

（1）初始化基数树。在系统初始化阶段，需构建基数树以支持服务标识的存储与匹配。首先，创建基数树的根节点作为树的起始点。随后，针对每个服务标识，依据分隔符将其拆解为层级路径，并按照路径层次依次插入到基数树中。在插入过程中，每个节点需存储关键信息，包括当前路径片段、子路径与子节点之间的映射关系，以及与服务相关的元数据，从而为后续的流量调度提供数据支撑。

（2）流量调度流程。当系统接收到客户端请求时，首先从请求中提取服务标识。接着，利用基数树进行服务匹配，从根节点开始，按照路径层级逐层查找。若当前路径片段在子节点映射中存在，则进入对应的子节点继续匹配；若路径片段不存在，则返回“未找到服务”的错误信息。一旦到达目标节点，根据节点元数据中的服务实例列表以及预设的负载均衡策略选择合适的服务实例。常见的负载均衡策略包括权重轮询、最小连接数以及健康检查。最后，将请求转发至选定的服务实例，完成流量调度过程。

为确保系统能够适应服务实例的动态变化，基数树提供了动态更新机制。当服务实例发生变更时，系统会及时更新对应节点的元数据，以反映最新的服务实例状态。若服务标识发生变化，系统会删除旧路径节点，并插入新的路径节点，确保服务标识的准确性。此外，系统还支持动态调整节点元数据中的权重值，调整后的权重会实时生效，进而影响流量分配。同时，系统会定期对服务实例进行健康检查，根据检查结果更新节点元数据中的健康标志，以便在流量调度时跳过不健康的实例，保证服务的可用性和稳定性。

以统一接入平台的白名单URL请求调度转发为例，白名单库超过2 000个，网站和小程序的白名单部分URL构建的白名单树如图7所示，当网站访问URL为/otn/login/pwd/时，统一接入平台会按照树匹配，先匹配到otn，继而匹配到login登录业务，最后匹配到账号密码登录，URL正确转发到登录服务的负载地址，若采用传统的正则表达式逐条匹配，每条规则需完整匹配整个URL路径，前缀/otn会被重复匹配。而基数树通过共享前缀节点，仅需匹配一次/otn即可在该分支下继续匹配后续路径，避免了冗余比较，显著提升了匹配效率。

## 3　实验效果分析

### 3.1　跨中心流量调度

针对基于神经网络的跨中心流量调度算法，其实验验证结果如下。

（1）实验验证神经网络训练。选取2025年3月1—31日的数据，每1 min采集一个样本，共计44 640条时序样本，损失函数采用均方误差(Mean Squared Error，MSE)。均方误差用于计算预测值与真实值之间误差平方的平均值，能反映预测值的整体偏离程度，对异常值比较敏感。迭代300次，函数达到收敛状态，随训练轮次变化的损失函数收敛曲线如图8所示。

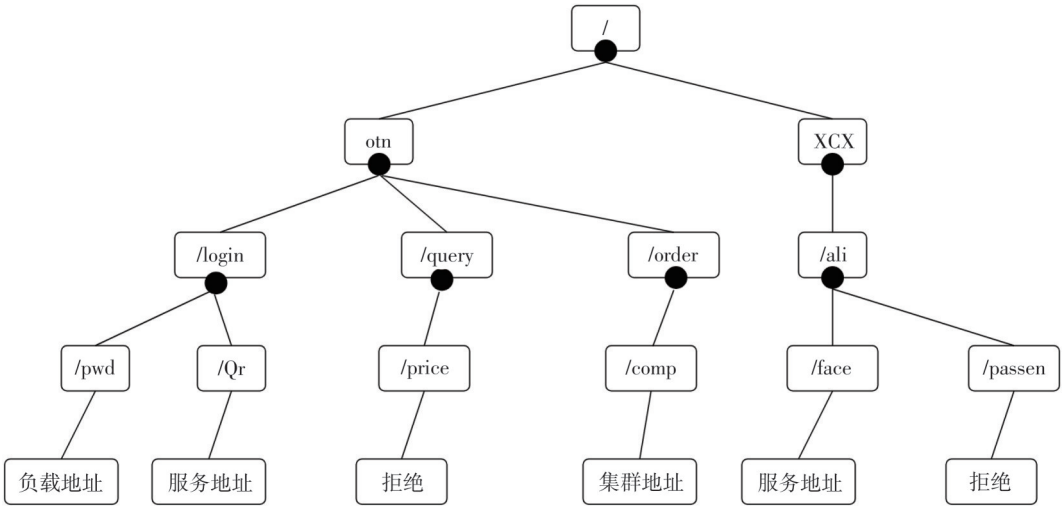


**图7　白名单树**

Fig.7　White list tree

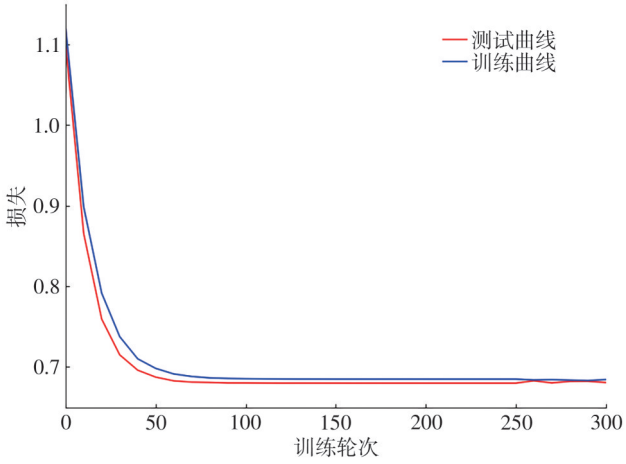


**图8　损失函数收敛曲线**

Fig.8　Convergence curve of loss function

（2）基于神经网络的流量调度学习。步长选择0.01，响应时长与调度比例的关系如图9所示，实验结果上看，响应时长随调度比例呈先下降后上升趋势，当调度比例为0.4时，响应时长最短达到1.8 s。实际应用时，可以每隔一段时间对流量分配进行一次寻优，从而使得系统处于最佳的运行状态。

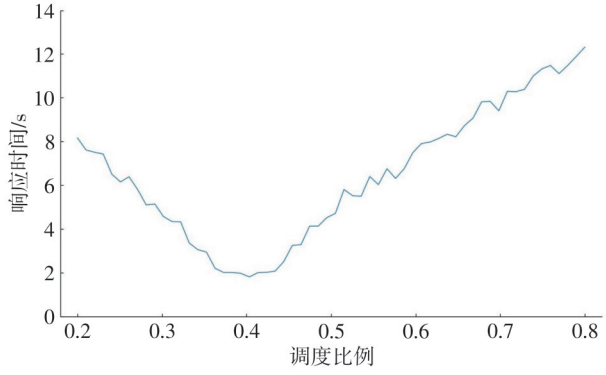


**图9　响应时长与调度比例的关系**

Fig.9　Relationship between response time and scheduling ratio

在测试环境中，按照模型推荐的最优比例进行调度，并与历史策略进行对比测试，测量实际响应时间、每秒请求数(Queries Per Second，QPS)等关键指标，不同策略的性能对比数据如表1所示。测试结果表明：相较于历史策略，基于神经网络的调度策略实际响应时间缩短了约15.5%，QPS提升了约16.1%，整体性能表现更优，为业务处理带来更快捷的响应。

**表1　不同策略性能对比数据**

Tab.1　Performance comparison of different strategies

| 策略 | 调度比例 | 响应时间/s | QPS/(万次/s) |
|---|---|---|---|
| 历史策略 | 0.5:0.5 | 2.13 | 8.7 |
| 神经网络 | 0.6:0.4 | 1.80 | 10.1 |

为平衡实时性与计算开销，调度周期设置为5 min。该间隔既可响应系统状态变化，又避免频繁计算对系统性能的影响。此外，为适应系统长期变化，模型支持定期重训练，以保持预测准确性。

## 3.2　数据中心内部流量调度

数据中心内部的流量调度以基数树为基础，能够实现高效长路径匹配，快速响应服务请求。实验硬件环境配置为CPU 128C ARM，内存1 TB，硬盘10 TB NVME，基于808个路由规则、URL平均长度150个字符串测试数据集，按照持续5 min并发2 000压测，基于正则表达式的匹配算法每秒事务数

(Transactions Per Second，TPS)为2 663、响应时间(Response Time，RT)为557 ms，基于基数树算法版本支持TPS为3 000，RT为400 ms，整体TPS性能提升12.7%，RT降低28.2%。

本研究设计了一套智能流量调度系统，涵盖跨中心调度与数据中心内部流量调度，在兼顾当前应用现状的同时，也为未来双CDN架构提供了扩展支持。该框架已在实践中取得显著成效。在内容分发层，基于CDN的两种双活调度方案实现区域级流量精准分配，降低了跨运营商访问延迟；通过统一接入平台，系统可实现双活数据中心快速切换，并且研发运维平台能够实现故障自动发现、隔离与恢复，利用订单集群、余票查询业务进行故障演练验证，可将故障平均修复时间降低至1 s。在流量调度问题上，通过是否启用风险流量过滤策略验证了动态路由策略拒绝非法请求的功能，启动策略时能够有效地降低异常访问量提升系统稳定性；通过在非售票期人为改变后端服务的流量分配比例，验证了资源的自动扩充缩减功能，提升了利用率，并且使用加权轮询、一致性哈希等策略，进一步优化了资源分配方案。综上，设计的智能流量调度系统有力保障了互联网售票系统核心业务的稳定运行，为旅客提供了更加流畅、便捷的购票体验，全面提升了系统可用性与用户满意度。

## 4　结束语

为铁路互联网售票系统构建了一套基于神经网络的智能流量调度解决方案，在架构层面，通过整合CDN双活与统一接入平台，实现了流量的灵活调度与精细管控；在算法层面，提出基于神经网络的流量预测模型和特定服务流量调度算法，提升了系统流量调度的精准性和实效性。该方案在春运等高并发场景中成功应用，显著提升了系统性能与资源利用率，有力提升旅客出行的购票体验。下一步，研究将致力于构建一个闭环的智能调度体系，通过实时学习机制精准感知网络状态与流量特征，输出双CDN流量分配的精准比例，并融合异常流量识别技术进行流量过滤，从而形成自动化、智能化的流量调度系统。

**参考文献：**

[1] 刘　俊. 铁路运输调度现代化发展策略研究[J]. 铁道运输与经济，2025，47(2)：1-5，24.
LIU Jun. Development Strategies of Railway Transportation Dispatching Modernization[J]. Railway Transport and Economy，2025，47(2)：1-5，24.

[2] 张克云，苗　茁，刘　岩，等. 面向发达高速铁路网的调度管理优化策略研究[J]. 铁道运输与经济，2024，46(9)：170-177，227.
ZHANG Keyun，MIAO Zhuo，LIU Yan，et al. Research on Dispatching Management Optimization Strategy for Developed High Speed Railway Network[J]. Railway Transport and Economy，2024，46(9)：170-177，227.

[3] 金子龙，钱　亮，赵　晨，等. 智能网络中基于流量预测的加权公平调度算法[J]. 电子测量技术，2024，47(18)：38-46.
JIN Zilong，QIAN Liang，ZHAO Chen，et al. Improved WFQ Algorithm Combined with Traffic Prediction in the Intelligent Network[J]. Electronic Measurement Technology，2024，47(18)：38-46.

[4] 郑　凌，魏国栋，潘伟涛，等. 面向车载时间敏感网络的时间感知最大匹配流量调度算法[J]. 物联网学报，2025，9(3)：48-59.
ZHENG Ling，WEI Guodong，PAN Weitao，et al. Time-Aware Traffic Scheduling Algorithm Based on Maximum Matching for In-Vehicle Time-Sensitive Networks[J]. Chinese Journal on Internet of Things，2025，9(3)：48-59.

[5] 高明明，郭孜政，骆泳吉，等. 朔黄铁路调度行车指挥智算平台总体架构研究[J]. 铁道运输与经济，2025，47(6)：109-118.
GAO Mingming，GUO Zizheng，LUO Yongji，et al. Overall Architecture of Intelligent Computing Platform for Shuozhou-Huanghua Railway Traffic Dispatching Command[J]. Railway Transport and Economy，2025，47(6)：109-118.

[6] 雷其鸣，牛　庚，桑学锋，等. 基于改进强化学习算法的水库群多目标优化调度[J]. 南水北调与水利科技(中英文)，2025，23(5)：1185-1195.
LEI Qiming，NIU Geng，SANG Xuefeng，et al. Multi-Objective Optimal Scheduling of Reservoir Group Based on Improved Reinforcement Learning Algorithm[J]. South-to-North Water Transfers and Water Science & Technology，2025，23(5)：1185-1195.

[7] 陆以勤，黄成海，陈嘉睿，等. 基于遗传算法的时间敏感网络调度方法[J]. 华南理工大学学报(自然科学版)，2024，52(2)：1-12.
LU Yiqin，HUANG Chenghai，CHEN Jiarui，et al. Time Sensitive Network Scheduling Method Based on Genetic Algorithm[J]. Journal of South China University of Technology (Natural Science Edition)，2024，52(2)：1-12.

[8] 黄宗伟. 基于SDN的数据中心网络大象流负载均衡调度策略研究[J]. 重庆科技学院学报(自然科学版)，2023，25(2)：76-81.
HUANG Zongwei. Research on Scheduling Strategy of Elephant Flow Load Balancing in Data Center Network Based on SDN[J]. Journal of Chongqing University of Science and Technology (Natural Sciences Edition)，2023，25(2)：76-81.

[9] 陈　杰，王俊昌，胡嘉晨，等. 面向多样化网络需求的优先级队列流量调度算法[J]. 计算机工程，2025，51(11)：152-161.
CHEN Jie，WANG Junchang，HU Jiachen，et al. Priority Queues-Based Traffic Scheduling Algorithm for Diverse Network Requirements[J]. Computer Engineering，2025，51(11)：152-161.

[10] 崔羽飞，毋　涛，刘　畅，等. 面向天地一体场景的异构算力调度与流量卸载应用研究[J]. 信息通信技术，2024，18(5)：48-53，70.
CUI Yufei，WU Tao，LIU Chang，et al. Research on Heterogeneous Computing Power Scheduling and Traffic Offloading Applications for Earth-Space Integrated Scenarios[J]. Information and Communications Technologies，2024，18(5)：48-53，70.

[11] 褚世冲，夏　栋，王　悦，等. 基于移动互联网的智能电梯调度系统[J]. 电子技术与软件工程，2016(17)：26-27.

[12] 代丽伟. 面向SDN的智能流量调度技术研究[D]. 沈阳：沈阳理工大学，2023.

[13] NIE H R，SU Y，ZHAO W B，et al. Hybrid Traffic Scheduling in Time-Sensitive Networking for the Support of Automotive Applications[J]. IET Communications，2024，18(2)：111-128.

[14] LI Q，LI D，JIN X，et al. A Simple and Efficient Time-Sensitive Networking Traffic Scheduling Method for Industrial Scenarios[J]. Electronics，2020，9(12)：2131.

[15] 徐　晶，刘春龙，霍佳皓，等. 基于可满足性模理论的时间敏感网络流量调度机制[J]. 计算机科学，2025，52(S2)：700-705.

[16] 李孟男. 面向数据中心网络的动态流智能调度技术研究[D]. 北京：北京邮电大学，2021.

[17] WAN J L，LIN S，ZHANG Z F，et al. Scheduling Real-Time Wireless Traffic：A Network-Aided Offline Reinforcement Learning Approach[J]. IEEE Internet of Things Journal，2023，10(24)：22331-22340.

(责任编辑　金　晶)