Scholar Publishing Group
International Journal of Multimedia Computing
https://doi.org/10.38007/IJMC.2025.060116
ISSN 2789-7168 Vol. 6, Issue 1:160-167

Research on Optimizing Advertising Service Stability under Peak Traffic

Jingtian Zhang

Georgia Institute of Technology, Atlanta 30332, Georgia, USA

Keywords: Peak traffic; Advertising services; Load balancing; disaster recovery

Abstract: This article discusses strategies for improving the stability of advertising services during peak traffic periods. Faced with the surge in advertising demand, advertising service services have encountered problems such as increased system carrying pressure, slowed response speed, and uneven resource allocation. Proposed strategies such as optimizing load balancing and traffic allocation, achieving elastic expansion and reasonable allocation of resources, conducting traffic prediction to optimize scheduling strategies, and strengthening system redundancy design and disaster recovery backup, with the aim of enhancing the stability and service quality of advertising service systems during peak traffic periods. Adopting these strategies can effectively handle traffic fluctuations, ensure the stability of advertising services, and protect the interests and returns of advertisers.

1. Introduction

In the context of the rapid growth of Internet advertising, ensuring the stability of advertising services is the core of the long-term operation of advertising services. Especially during peak hours, advertising services face the pressure of large-scale concurrent requests and sudden traffic, with frequent occurrences of system overload, response delays, and improper resource scheduling, seriously affecting the effectiveness of advertising services and user experience. Therefore, ensuring the smooth operation of advertising services in high traffic environments has become a key issue that advertising operators urgently need to address. This article will delve into the various challenges that peak access brings to advertising services, and provide improvement plans with the aim of enhancing the stability and service level of advertising operations during peak access periods, ensuring the accuracy and efficiency of advertising releases.

2. Overview of advertising service stability

The robustness of advertising platforms is reflected in their ability to maintain service continuity and trustworthiness even in situations where user traffic increases dramatically and data traffic is massive. The aspects covered by stability include but are not limited to the following: firstly, the rationality of the system architecture. Advertising platforms need to handle massive ad requests, and during peak periods, the system may experience overload operations, resulting in service delays or interruptions. The second is the stability of resource allocation, which is closely related to fluctuations in advertising traffic. When there is a significant change in traffic, the rationality of resource allocation becomes particularly important[1]. If the system cannot flexibly adjust resources, it can easily cause resource congestion or waste, exacerbate system performance bottlenecks, and affect the normal display of advertisements. The third issue is the inadequate management of sudden traffic and changes in traffic. Traditional advertising platforms often lack flexible response strategies when facing sudden traffic growth, resulting in service interruptions or delays. This not only damages the user experience, but may also result in the advertiser's investment being wasted. To prevent such situations, advertising platforms need to have the ability to monitor traffic in real-time and dynamically adjust resources to cope with the challenges brought by changes in traffic. The fourth is the system's backup and disaster recovery capabilities. If the system lacks a sound backup design and disaster recovery plan, once a failure occurs, advertising services may not be able to recover quickly, resulting in long-term service interruptions. In order to enhance the system's ability to resist risks, advertising platforms need to build a multi-level backup system to ensure rapid switching in the event of a single point of failure, ensuring the continuity of services. In short, ensuring the stability of advertising services not only requires a reasonable system structure, efficient resource allocation, and traffic control, but also relies on a comprehensive backup and disaster recovery mechanism. Only by being prepared at these levels can advertising platforms maintain stable operation during peak traffic and ensure the accuracy of advertising placement.

3. The current situation of advertising services during peak traffic hours

3.1 System overload and response delay

During peak traffic periods, advertising services must cope with massive synchronous requests, especially during critical periods such as promotional seasons or holidays when user traffic sharply increases. The resource scalability of most advertising platforms is insufficient, resulting in heavy system load and a significant decrease in request processing speed. In order to quickly respond to the needs of advertisers and push advertising content in a timely manner[2], once the server cannot process the huge amount of data in a timely manner, advertising requests will be backlogged, causing a delay in response time. The increase in ad loading time, the confusion of displayed content, and even loading failures directly affect the advertiser's return on investment. At the user experience level, delayed response can cause inconvenience to users, thereby reducing their satisfaction. More seriously, system overload may result in some advertisements not being displayed properly, and in extreme cases, the entire advertising platform may crash, causing service interruptions and customer loss.

3.2 Improper resource scheduling and system bottlenecks

At the moment of a surge in traffic, advertising publishing systems must cope with massive synchronous requests and complex data processing tasks, which poses a severe challenge to the resource allocation efficiency of the system. However, many advertising publishing systems lack a well-designed resource allocation mechanism, resulting in imbalanced resource allocation during high load operations. Some key components, especially database and Internet resources, may become system performance bottlenecks due to unreasonable allocation. Advertising requests may encounter bandwidth shortage during network transmission, resulting in request backlog and increased latency. The database lacks the ability to handle a large number of read and write requests, resulting in slow data processing and exacerbating the pressure on the entire system. This unreasonable allocation of resources directly affects the stability of advertising display, leading to failed or inaccurate advertising placement. The poor quality of advertising display not only affects the effectiveness of advertising, but may also have a long-term impact on the reputation of advertisers and platforms.

3.3 Insufficient management of sudden traffic and traffic fluctuations

Advertising services have faced severe challenges in dealing with traffic fluctuations. Due to the platform's inability to accurately predict the scale and timing of traffic fluctuations, it is easy for traffic surges to exceed the platform's carrying capacity, resulting in system overload, delayed or failed advertising display[3]. Many advertising platforms have not yet implemented real-time traffic adjustment mechanisms, which cannot automatically follow traffic changes to adjust system load, resulting in ineffective allocation of computing resources during peak traffic periods. Although some services attempt to utilize load balancing technology to distribute traffic, when faced with rapid traffic growth, the system often struggles to respond in a timely manner, unable to quickly capture changes in traffic fluctuations and take corresponding measures. Faced with increasingly diverse industry demands, advertising services urgently need more flexible and intelligent traffic management strategies to adapt to resource allocation under different traffic conditions. However, many current systems often fall into management chaos during peak traffic periods due to the lack of efficient traffic prediction methods.

3.4 Lack of redundancy design and disaster recovery capability

Redundancy design and disaster recovery capability are the foundation for ensuring the continuous and stable operation of advertising services. However, the construction of many advertising services in this critical field is not sound. During periods of surge in traffic, if the core components of the advertising system are not designed with backup, any single component failure may cause the entire service to shut down. For example[4], if key facilities such as advertising service servers, databases, or load balancers do not have backup mechanisms, once a failure occurs, they cannot be smoothly switched to the backup system, which in turn affects the normal display of advertisements. In this case, the service interruption period of advertising services is usually long and the recovery speed is slow, which directly weakens the effectiveness of advertising services. The lack of effective disaster recovery strategies and mechanisms also makes it difficult for advertising service services to quickly restore normal operation in emergency situations such as server crashes or hardware damage.

4. Optimization plan for advertising service stability during peak traffic periods

4.1 Load balancing and traffic scheduling optimization

Advertising services require the establishment of an intelligent load balancing system. Track user requests and server load status through a real-time traffic prediction system. Once there is a significant change in traffic, the system estimates the peak traffic period and its potential fluctuation range based on past data and user behavior patterns. When implementing load balancing strategies, services also need to optimize traffic allocation schemes accurately based on factors such as the type of advertising request, user geographic location information, device category, etc. Advertising platforms use geographic location information to direct user requests to the nearest server node, thereby reducing network latency and enhancing user interaction experience. The platform can also use multi-level load balancing strategies[5], including global load balancing and local load balancing, to ensure even distribution of network traffic. The advertising publishing system has greatly improved its adaptability and scalability by adopting containerization technology and microservice architecture. By utilizing containerization techniques, the system can automatically add service entities to alleviate traffic pressure during sudden increases in traffic. The microservice architecture divides the various functional units of the advertising system into multiple autonomous service units, allowing each unit to independently expand as needed, solving the problem of centralized pressure on all functional modules under traditional single architecture. As shown in Table 1, the core strategies for achieving load balancing and traffic scheduling include traffic estimation, traffic distribution based on geographic location, multi-level load balancing, and the comprehensive use of containerization and microservices. These measures work together to ensure the smooth operation of the advertising publishing system during peak traffic periods.

Table 1. Key Measures for Load Balancing and Traffic Scheduling Optimization

Optimization measures	Implementation steps	Role
Traffic prediction and monitoring	Based on historical data and user behavior patterns, predict peak traffic and fluctuation amplitude.	Identify traffic fluctuations in advance and prepare for resource scheduling.
Geographic based traffic allocation	Based on the user's geographic location and request type, direct traffic to the nearest server node.	Reduce network latency and enhance user experience.
Multi level load balancing	Optimize traffic allocation by combining global load balancing and local load balancing	Avoid single point overload and ensure even distribution of traffic
Containerization and microservice architecture	Using containerization technology to expand service instances and adopting microservice architecture to share load	Enhance system flexibility and avoid system overload

4.2 Elastic Expansion and Resource Allocation Adjustment

Elastic scaling technology can flexibly increase computing resources during peak traffic periods through automated resource adjustments, ensuring the continuous availability of advertising services[6]. This service relies on real-time monitoring mechanisms to continuously track the resource usage status of server nodes, covering key indicators such as central processor utilization, memory usage, and network transmission bandwidth. If the load of a node exceeds the predetermined limit, the system will automatically activate the expansion process and start a new server instance to share the pressure. To this end, the platform first sets a series of thresholds and rules, such as automatically adding computing instances to balance the load once the node load reaches the peak threshold. Advertising platforms also need to allocate resources accurately, especially in environments with significant fluctuations in resource demand. For key business units such as advertising service systems, data analysis modules, and user request processing centers, service priority is given to ensuring that they have sufficient computing resources to ensure efficient operation during high traffic periods. In order to further optimize resource allocation, the platform can also prioritize the needs of core customers and key advertisers by setting resource priorities. This strategy can be implemented through a priority scheduling system, which can flexibly adjust the resource allocation shares of each advertiser based on their financial budget, expected advertising effectiveness, and real-time traffic changes. At the same time, the service also adopts microservice architecture and container technology to enhance the scalability of the system.

Through microservice architecture, the advertising system is subdivided into numerous independent business units, and the resource utilization of each unit can be separately allocated and expanded. As shown in Table 2, advertising services can achieve flexible elastic expansion and efficient resource management through setting thresholds, priority resource allocation, and microservice architecture.

Table 2. Elastic Expansion and Resource Allocation Adjustment Strategies

strategy	implementation steps	effect
Load monitoring and threshold setting	Continuously monitor node load and trigger expansion mechanism when load exceeds threshold	Ensure fast expansion of computing resources during peak traffic periods
Prioritize resource allocation for core modules	Prioritize resource allocation for core modules such as advertising services and data analysis	Ensure that critical business operations remain efficient even during traffic fluctuations
Priority resource scheduling	Adjust resource allocation based on advertiser budget, goals, and real-time traffic demands	Prioritize the effectiveness of advertising services for high-value customers
Microservices and Containerized Architecture	Using microservice architecture to split functional modules and containerization technology to achieve resource expansion	Improve system flexibility, reduce resource waste, and respond quickly

4.3 Traffic prediction and scheduling optimization strategies

Optimizing traffic estimation and scheduling strategies plays a crucial role in maximizing resource utilization in advertising services and ensuring the stability of advertising releases. The service must rely on numerous factors such as historical statistical data, user behavior pattern analysis, and seasonal changes to create an accurate traffic estimation model. In daily operation[7], the service needs to summarize and deeply analyze the traffic information of past advertising services, extract periodic characteristics from it, such as peak traffic time, traffic change patterns, etc., and then estimate traffic based on these information. After the estimation work is completed, the service needs to build an intelligent resource scheduling system based on the estimated data. By utilizing massive data analysis, services can allocate resources reasonably according to the specific needs of different advertising clients. The service will independently optimize the resource allocation of the advertising service module based on the predicted peak traffic, and adjust the service tactics accordingly according to multiple factors such as the advertiser's capital investment and advertising effectiveness. In order to enhance the flexibility of traffic management, the system also needs to change the traffic allocation plan in real time. The system will monitor real-time advertising data and user activity to ensure timely response to traffic fluctuations. During peak traffic periods, the service utilizes advanced intelligent traffic management algorithms, supplemented by real-time data feedback, to flexibly adjust traffic allocation. In this process, the service relies on precise prediction and efficient intelligent regulation methods to ensure the rationality of resource allocation and the sustained stability of advertising services, thereby improving the advertising service level and user experience during traffic fluctuations.

4.4 Redundancy design and disaster recovery plan

Advertising platforms need to establish a multi-level redundant design system, in which the platform deploys multiple data centers in different geographical locations to form a cross regional redundant network. Through load balancing technology, the system intelligently distributes requests to different data centers, ensuring that in the event of a problem in any data center, requests can be automatically redirected to normally functioning nodes, thereby preventing service interruptions[8]. Within each data center, multiple backup servers and diverse data paths should also be deployed to avoid service disruptions caused by a single point of failure. The backup strategy should not only cover data storage, but also key areas such as computing power, network transmission, and databases. At the same time, disaster recovery plans should complement backup strategies. The advertising system needs to perform regular backup operations to ensure secure storage of data in multiple locations and prevent the risk of data loss. In order to better demonstrate redundancy design and disaster recovery solutions, the following is the redundancy design architecture diagram of the advertising platform:

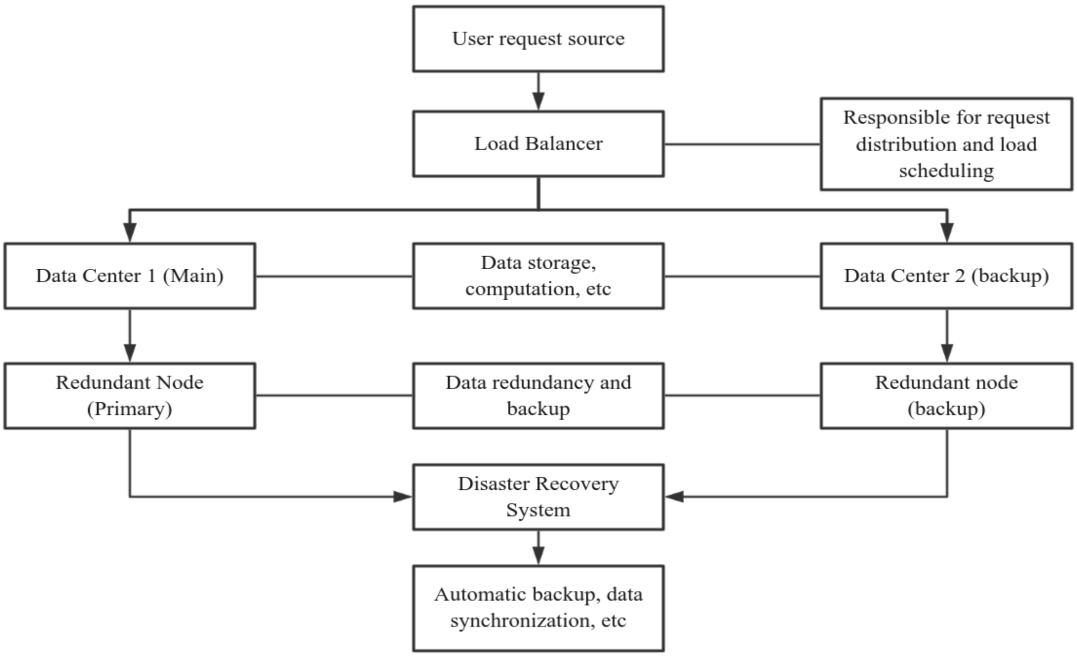
User request source
Load Balancer
Responsible for request distribution and load scheduling
Data Center 1 (Main)
Data storage, computation, etc
Data Center 2 (backup)
Redundant Node (Primary)
Data redundancy and backup
Redundant node (backup)
Disaster Recovery System
Automatic backup, data synchronization, etc

Figure 1. Redundancy Design and Disaster Recovery Architecture Diagram

The system should also conduct regular disaster simulation exercises to verify the effectiveness of the recovery process. Once the system encounters a failure, the disaster recovery mechanism will be quickly activated according to established procedures, using automated tools to quickly recover damaged parts, striving to minimize service interruption time. The advertising system should also develop detailed troubleshooting procedures and emergency response plans. The technical team needs to follow established procedures for fault location, repair, and system recovery to ensure quick response and effective handling in the face of unexpected situations. With a carefully designed backup architecture and disaster recovery plan, the advertising system can ensure high availability and stability during peak traffic or unexpected situations, ensuring the continuity and quality of advertising services.

5. Conclusion

At the critical moment when user traffic surges, the stability of advertising services is facing significant challenges, such as system overload, delayed response, and improper resource allocation, all of which have a direct impact on the smooth operation and advertising effectiveness of the advertising system. Therefore, the strategies proposed in this study, such as balancing load, flexible expansion, predicting traffic, and setting redundancy, provide strong support for the smooth operation of advertising systems during peak user access periods. By adopting intelligent traffic allocation, dynamic resource adjustment, and a sound fault recovery system, services can more easily cope with sudden changes in traffic, ensuring the accuracy and efficiency of advertising services. In the future, with the improvement of technological level, advertising services need to continuously improve their system architecture and resource management methods to adapt to the constantly changing traffic situation.

Reference

[1] *Thanh B K. The role of self-efficacy and firm size in the online advertising services continuous adoption intention: Theory of planned behavior approach. Journal of Open Innovation: Technology, Market, and Complexity, 2023, 9(1).*

[2] *Yu Y, Sheng C. A Study on Data Monitoring and Effect Optimization of Programmed Advertising Platform: Taking “Ocean Engine” as an Example. Journal of Physics: Conference Series, 2021, 1883(1).*

[3] *Yong H, Yanan Y, Zhongyuan W, et al. Equilibrium Pricing, Advertising, and Quality Strategies in a Platform Service Supply Chain. Asia-Pacific Journal of Operational Research, 2022, 39(1).*

[4] *Joel R, B. N M, Vassilis B, et al. A new model for optimal advertising impression allocation across consumer segments. Applied Marketing Analytics, 2023, 9(2):117-133.*

[5] *Utkarsh. Tangible and intangible quality cues in service advertising: A construal level theory perspective. Journal of Global Scholars of Marketing Science, 2023, 33(1):90-106.*

[6] *Gao B, Wang Y, Xie H, et al. Artificial Intelligence in Advertising: Advancements, Challenges, and Ethical Considerations in Targeting, Personalization, Content Creation, and Ad Optimization. SAGE Open, 2023, 13(4)*

[7] *Yang Y, Feng B, Salminen J, et al. Optimal advertising for a generalized Vidale–Wolfe response model. Electronic Commerce Research, 2021, 22(4):1-31.*

[8] *Y. Zhao, "Design and Financial Risk Control Application of Credit Scoring Card Model Based on XGBoost and CatBoost, " 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-5.*

[9] *B. Li, "Research on the Spatial Durbin Model Based on Big Data and Machine Learning for Predicting and Evaluating the Carbon Reduction Potential of Clean Energy," 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-5,.*

[10] *Q. Xu, "Implementation of Intelligent Chatbot Model for Social Media Based on the Combination of Retrieval and Generation,” 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 2025, pp. 1-7.*

[11] *Su H, Luo W, Mehdad Y, et al. Llm-friendly knowledge representation for customer support[C]//Proceedings of the 31st International Conference on Computational Linguistics: Industry Track. 2025: 496-504.*

[12] *F. Liu, "Architecture and Algorithm Optimization of Realtime User Behavior Analysis System for Ecommerce Based on Distributed Stream Computing, " 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-84.*

[13] Y. Zou, "Research on the Construction and Optimization Algorithm of Cybersecurity Knowledge Graphs Combining Open Information Extraction with Graph Convolutional Networks, " 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 2025, pp. 1-5.

[14] M. Zhang, "Research on Joint Optimization Algorithm for Image Enhancement and Denoising Based on the Combination of Deep Learning and Variational Models, " 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-5.

[15] W. Han, "Using Spark Streaming Technology to Drive the Real-Time Construction and Improvement of the Credit Rating System for Financial Customers, " 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-6.

[16] Ding, J. (2025). Research On CODP Localization Decision Model Of Automotive Supply Chain Based On Delayed Manufacturing Strategy. arXiv preprint arXiv:2511. 05899.

[17] Wu Y. Optimization of Generative AI Intelligent Interaction System Based on Adversarial Attack Defense and Content Controllable Generation. 2025.

[18] Wang Y. Application of Data Completion and Full Lifecycle Cost Optimization Integrating Artificial Intelligence in Supply Chain. 2025.

[19] Chen M. Research on Automated Risk Detection Methods in Machine Learning Integrating Privacy Computing. 2025.

[20] Wei, X. (2025). Deployment of Natural Language Processing Technology as a Service and Front-End Visualization. International Journal of Engineering Advances, 2(3), 117-123.