By Olivia Wilson

Server Pressure Due To High Traffic Demand

research paper

Alakeel, A. M. (2010). "Efficient Load Balancing Algorithms: Comparative Study."

This paper provides a comparative analysis of various load balancing algorithms used in distributed systems. The study highlights the importance of efficient load balancing in reducing server strain, particularly during periods of high traffic. The paper examines different algorithms based on parameters like response time, throughput, and resource utilization.

Alzoubi, H., Alsmadi, I., and Hsu, C.-H. (2021). "Traffic Load Balancing in Cloud Data Centers Using Dynamic Resource Allocation."

- This research focuses on traffic load balancing within cloud data centers, emphasizing dynamic resource allocation. The paper explores how cloud systems manage traffic and balance loads to prevent bottlenecks, ensuring optimal performance even during peak traffic periods. It introduces techniques to improve server responsiveness and reduce downtime by reallocating resources dynamically.

Prusty, N. (2018). "Blockchain for Enterprise: Build Scalable Blockchain Applications with Privacy, Interoperability, and Permissioned Features."

This book delves into the application of blockchain technology in enterprise environments, emphasizing scalability, privacy, and security. While not directly related to load balancing, it explores the challenges of building large-scale applications in a decentralized system and the methods of ensuring efficient use of resources.

Problem statement

Even with methods like load balancing, caching, and auto-scaling, servers still struggle during sudden traffic surges. Our research explores these issues and potential improvements.

Agenda

01. Key Objectives
02. Methods
03. Results

01

Objectives

Objective 1

Identify the causes of server pressure

Objective 2

Review current solutions.

Objective 3

Suggest enhancements for better traffic management.

02

methods

01 Load Balancing: Distributes traffic across servers.

02 Auto-Scaling: Adjusts the number of servers based on demand.

03 Caching: Stores frequently used data to reduce server load.

04 CDNs: Reduces latency by distributing content closer to users.

03

result

first

Load Balancing reduced response times by 30-50%, but some algorithms didn't always handle uneven loads well.

second

Auto-Scaling improved response times by 45%, but had latency issues during sudden spikes.

third

Caching reduced server load by up to 50% in high-traffic situations.

Future works

- Proactive Traffic Management:
- Develop proactive traffic management strategies using predictive techniques like machine learning and AI to anticipate traffic surges and scale server resources preemptively.
- Cost-Effective Solutions for Smaller Enterprises:
- Investigate affordable alternatives for dynamic resource scaling that cater to small and medium-sized businesses without incurring high operational costs.
- Advanced Caching Mechanisms for Dynamic Content:
- Develop enhanced caching mechanisms for real-time or frequently updated data, useful for platforms like social media or stock trading.
- Refinement of Load Balancing Algorithms:

Improve load balancing algorithms by considering factors like geographic location, server latency, and user session persistence to optimize traffic distribution.