AI & CYBER FORUM
AN INTERNATIONAL JOURNAL
OPEN ACCESS
ISSN 3117-6399
Original Research Article
Check for updates

Dynamic Markov Model for Resource Contention Prediction in an Edge Computing Architecture

Madhukumar K* and **Prerna Sidana**

Glocal School of Technology & Computer Science, Glocal University, Saharanpur, Uttar Pradesh, India

Corresponding author: **Madhukumar K** | E-mail: **madhukumarkreddy@gmail.com**

Citation: Madhukumar K and Prerna Sidana (2025). Dynamic Markov Model for Resource Contention Prediction in an Edge Computing Architecture. *AI & Cyber Forum: An International Journal.* **DOI: https://doi.org/10.51470/AI.2025.4.2.18**

Received 15 June 2025 | Revised 13 July 2025 | Accepted 25 July 2025 | Available Online August 09, 2025

Abstract

Edge computing seeks to minimize processing and storage demands on the cloud while delivering services with decreased latency. Devices near the network's periphery include smartphones, tablets, routers, surveillance footage, presses, and other IoT gadgets that generate copious amounts of data. Virtual machines (VMs) located on edge servers close to the edge devices handle the data. In order to anticipate power congestion at edge computers, this research presented two models: the Dynamic Markov Chain with Regression and Markov (DMRCP) and the Hybrid Cascading of Reconstruction and Markov (CRMRCP). In order to evaluate performance, the research examines the amount of virtual machine migrations and compares the DMRCP method with initial-order and second-degree Markov models. We compare the number of virtual machine deployments using the CRMRCP method to that using the third-order Poisson model. Based on the findings, the DMRCP requires fewer virtual machine migrations compared to the first-order equations Markov algorithm (52.9%) and the second-degree Population model (21.1%). By combining ridge regression with the second-degree Markov model, the number of virtual machine migrations in CRMRCP is reduced by 81.8%.

Keywords: *Resources Contention; Prediction; Edge computing; Edge computing; Architecture*

INTRODUCTION

Instead of sending computing to faraway cloud data centers, edge computing moves a lot of the work to the network's perimeter. Phones, printing machines, gadgets, computers, tablets, and Internet of Things (IoT) gadgets make up edge devices, each serving various functions. These gadgets produce substantial quantities of data [1, 2]. Edge servers are situated near edge devices and execute computations, with only essential data transmitted to the cloud. This leads to decreased latency. The resources required to perform operations on edge servers may include physical resources such as storage devices, CPUs, or network devices, as well as virtual resources like virtual machines (VMs). IoT workloads frequently experience fluctuations in resource requirements due to events at the edge server that necessitate more resources for the intended computation [3, 4]. Edge servers possess constrained hardware resources in comparison to cloud servers, resulting in heightened resource contention concerns. Consequently, each edge server can support only a certain number of VMs, rendering the allocation of VMs in an edge computing environment a complex endeavour [5]. The major purpose of this study is to offer a "dynamic Markov model for resource contention prediction" in an edge computing architecture.

The next part provides a detailed overview of the previous literature that is relevant to this investigation.

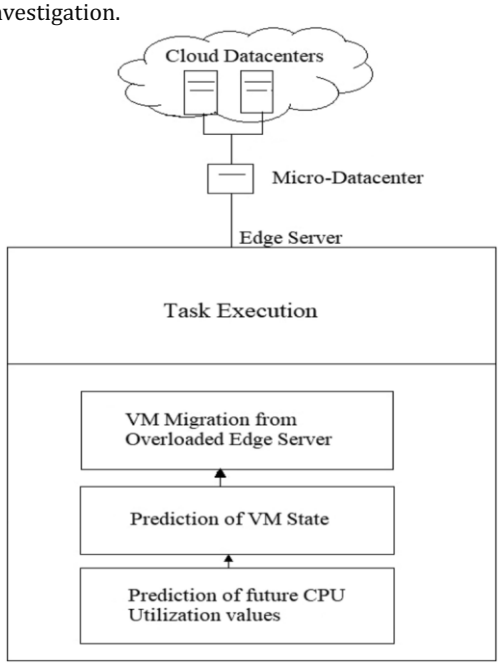

Figure 1: Prediction and migration of resources in edge computing [6]

© ***Authors:*** *Published in AI & Cyber Forum: An International Journal under the CC BY-NC-ND 4.0 license (http://creativecommons.org/licenses/by-nc-nd/4.0/). No commercial use or modifications permitted.*

Literature review

The following part provides a more in-depth discussion of the previous research that is relevant to this study, which is "Dynamic Markov model for resource contention prediction in an edge computing architecture."

Table 1: Related works

Authors and year	Methodlogy	Findings
Iftikhar et al., (2023) [7]	This SLR used traditional review methods to look into artificial intelligence and machine learning algorithms and the issues they cause with the administration of resources in edge computing and fog computing environments. Additionally, topics such as edge AI control and artificial learning/deep learning/reinforcement learning have been covered.	Open difficulties and intriguing future research directions in AI/ML-based fog/edge computing have been identified and discussed.
Tuli et al., (2023) [8]	This article explores how AI might enhance conventional applications for quality of service within a wide range of resources and dissects new AI approaches for managing services on the edge, in the fog, and in the cloud.	Presented some far-out concepts and saw this effort as laying the groundwork for future AI-powered computer system advancements.
Acheampong et al., (2023) [9]	The methods, strategies, and variables that influence task offloading as well as the allocation of resources in the edge of computing were thoroughly investigated in this paper.	The following methods were covered: GBD, Dynamic Voltage Scaling, Heuristic Algorithms Lyapunov Efficiency of Games Theory, and Non(convex) Optimized.
Golec et al., (2024) [10]	The study introduced MASTER, a machine learning-based resource management framework that predicts cold start delay for Industry 4.0 applications using an extreme gradient boosting (XGBoost) model for performance optimization.	Considering research challenges and limitations in computer science at the edge, this referenced paper concluded.

Research Gap: Due to their dynamic and decentralized character, edge computing architectures make resource contention prediction difficult. Static or simplistic dynamic models that fail to reflect real-time resource demand and availability dominate current research. A Dynamic Markov Model (DMM) could close this gap. Using real-time edge resource usage trends, a DMM might adapt transition probabilities, unlike stationary Markov models. This adaptive capacity would improve contention forecasts, enabling proactive resource allocation and management in edge computing environments, an understudied area.

Methodology

A qualitative research approach with secondary data collecting is used to investigate edge computing resource contention prediction literature and propose a Dynamic Markov Model (DMM) as an enhanced predictive framework. Data comes from peer-reviewed academic articles, conference proceedings, white papers, and technical reports on edge computing, Markov-based predictive models, and resource management. The study analyses existing models, finds holes in their capacity to react to real-time variations, and synthesizes information to support a DMM. The proposed approach dynamically adjusts transition probabilities based on real-time resource usage trends to improve contention prediction. Critical analysis of the literature develops the DMM's conceptual foundation, making it a revolutionary edge computing resource management tool.

Results and Discussions

This section discusses the settings-based evaluation outcomes. A surge in peripheral cloud virtual computer motions hurts application operation. As indicated in the graphic, the median number of virtual machine moves for the primary sequence, second order, and suggested Dynamical Markov Model (DMM) in Planetlab workload occurs over ten days. The first class of Markov model undergoes many virtual device transfers during the 10-day Planetlab workload. The third-order Markov model averages 27.2% fewer Vmigrations than the previous framework.

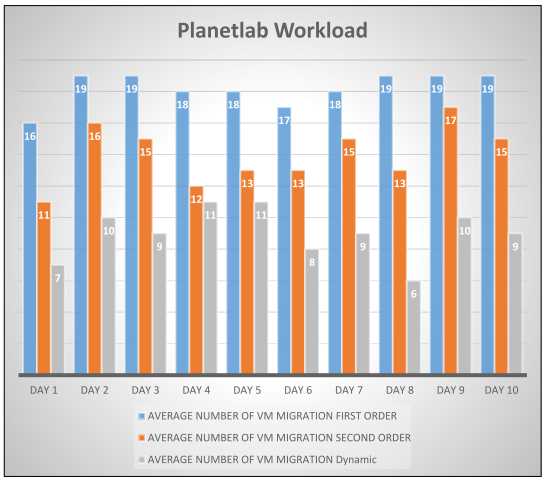

Figure 2: Planetlab Workload

In this section, a graphical analysis of the results gained from the experiments is presented.

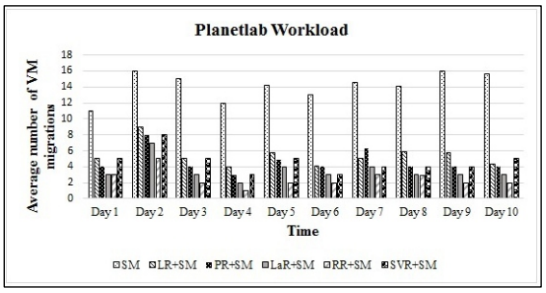

Figure 3: "Comparison of Number of VM Migrations in the CRMRCP Model with SM Model"

Results show that when the following level Markov model is employed alone, more virtual computers migrate than when reversion is additionally used. Logistic regression plus third-order Markov theory (LR+SM) reduce remote system migrations by 62.9% compared to SM. Third-order Markov theory (PR+SM) with regression waterfall polynomials reduce virtual computer journeys by 67.1%. Modeling the system with the subsequent order Markov system (LaR+SM) and chain of lasso regression reduces virtual computer migration by 75.2% compared to SM.

Compared to the SM, the SVR+SM cascade of regression algorithms reduces virtual computer motions by 68.3%. Ridge modeling and the subsequent order Markov algorithm (RR+SM) provide the fewest operating system transfers of all regression methods. The standard number of virtual computer transfers drops to 81.8%, especially against SM.

Conclusion

This research points out the shortcomings of current resource contention prediction models in edge computing and suggests a Dynamic Markov Model (DMM) as a more flexible alternative. The research uses secondary data to support the requirement for real-time transition probability adjustments, which improves forecast accuracy and optimizes resource management in dynamic edge computing systems.

References

1. Samir, A., & Pahl, C. (2019, June). Detecting and predicting anomalies for edge cluster environments using hidden markov models. In *2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 21-28). IEEE.
2. Masdari, M., & Khezri, H. (2020). Efficient offloading schemes using Markovian models: a literature review. *Computing*, 102(7), 1673-1716.
3. Ergun, K., Ayoub, R., Mercati, P., & Rosing, T. Š. (2022). Dynamic Reliability Management of Multigateway IoT Edge Computing Systems. *IEEE Internet of Things Journal*, 10(5), 3864-3889.
4. Duc, T. L., Leiva, R. G., Casari, P., & Östberg, P. O. (2019). Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey. *ACM Computing Surveys*, 52(5), 1-39.
5. Zakarya, M., Gillam, L., Ali, H., Rahman, I. U., Salah, K., Khan, R., & Buyya, R. (2020). epcAware: A game-based, energy, performance and cost-efficient resource management technique for multi-access edge computing. *IEEE Transactions on Services Computing*, 15(3), 1634-1648.
6. Surya, K., & Rajam, V. M. A. (2023). Novel Approaches for Resource Management Across Edge Servers. *International Journal of Networked and Distributed Computing*, 11(1), 20-30.
7. Iftikhar, S., Gill, S. S., Song, C., Xu, M., Aslanpour, M. S., Toosi, A. N., & Uhlig, S. (2023). AI-based fog and edge computing: A systematic review, taxonomy and future directions. *Internet of Things*, 21, 100674.
8. Tuli, S., Mirhakimi, F., Pallewatta, S., Zawad, S., Casale, G., Javadi, B., & Jennings, N. R. (2023). AI augmented Edge and Fog computing: Trends and challenges. *Journal of Network and Computer Applications*, 216, 103648.
9. Acheampong, A., Zhang, Y., Xu, X., & Kumah, D. A. (2023). A review of the current task offloading algorithms, strategies and approach in edge computing systems. *Computer Modeling in Engineering & Sciences*, 134(1), 35-88.
10. Golec, M., Gill, S. S., Wu, H., Can, T. C., Golec, M., Cetinkaya, O., & Uhlig, S. (2024). Master: Machine learning-based cold start latency prediction framework in serverless edge computing environments for industry 4.0. *IEEE Journal of Selected Areas in Sensors*.