



A Multi-Criteria VM Consolidation Approach for Service Level Agreement Compliance and Energy Efficiency in Cloud Data-Centers

Farkhondeh Kiaee*¹

¹ Assistant Professor, Department of Electrical Engineering, National University of Skills (NUS), Tehran, Iran

ARTICLE INFO

Received: 29.06.2024

Revised: 02.12.2024

Accepted: 27.01.2025

Keyword:

Cloud Computing,
Technique for Order of
Preference by Similarity to Ideal
Solution (TOPSIS),
Principal Component Analysis,
Consolidation,
Resource Management

*Corresponding Author:

Farkhondeh Kiaee

Email: kiaei@shariaty.ac.ir

ABSTRACT

Consolidating Virtual Machines (VMs) is an effective method to improve the energy efficiency of cloud environments. By using live VM migration, multiple VMs can be consolidated onto a minimal set of physical resources, allowing the unused hosts to be powered down. However, VM consolidation should not lead to performance degradation and Service Level Agreement (SLA) violations. This study introduces an efficient VM selection method for resource consolidation in cloud environments. The proposed method combines principal component analysis (PCA) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), which help in extracting uncorrelated criteria effects and ranking the options. By leveraging PCA to eliminate dependency between criteria and TOPSIS for intelligent ranking, the method avoids the bias of traditional multi-criteria approaches towards alternatives that have good evaluations in two or more dependent criteria. Simulation results using the Cloudsim simulator demonstrated the method's effectiveness, showing reductions of up to 41.5% in energy consumption, 30.13% in SLA violations, 12.9% in response time, 13.4% in running costs, and 58.2% in VM migrations compared to state-of-the-art methods.

EXTENDED ABSTRACT

Introduction

Virtualization-based consolidation is a practical method to enhance energy efficiency in Cloud environments. By leveraging live migration of virtual machines (VMs) and containers, multiple VMs can be consolidated onto a minimal set of physical resources, allowing the remaining idle hosts to be put into sleep mode or powered down. However, managing resources in Cloud environments is challenging due to heterogeneous resources, fluctuating resource demands, and conflicting goals for consolidation.

Most current consolidation algorithms are simplistic, considering only a limited number of factors for VM migration selection. They often fail to optimize migration for both energy efficiency and performance in heterogeneous Cloud environments. Additionally, they focus on correlated decision criteria, leading to biases that favor VMs with strong evaluations across dependent criteria. These approaches overlook the need to extract uncorrelated contributions from decision criteria, which is a gap this study addresses.

This paper introduces an efficient multi-criteria VM consolidation solution for heterogeneous Cloud environments. The proposed method applies Principal Component Analysis (PCA) to isolate the uncorrelated contributions of decision criteria. Using the simplified decision matrix, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is then employed to rank and select the best VMs for migration.

Methodology

The typical VM consolidation problem in cloud datacenters consists of four key phases: identifying overloaded hosts, identifying underloaded hosts, selecting VMs for migration from over-utilized hosts, and determining new placements for migrating VMs. Effective resource allocation can reduce the number of required hosts, minimize energy consumption and costs, and decrease SLA violations.

The proposed framework utilizes the TOPSIS multi-criteria algorithm to address the third phase, which is VM selection for migration. The input criteria for TOPSIS include: (1) the resource correlation between a VM and the other VMs on its host, (2) the CPU utilization of the VM, (3) the computing capacity of the VM in MIPS, and (4) its storage capacity.

However, current TOPSIS-based methods often assume independence among the criteria, which may not hold true since the criteria originate from the same VM. This can lead to biased results, favoring VMs with high scores in correlated criteria. The proposed framework overcomes this by incorporating PCA, which transforms the decision criteria into an uncorrelated set before applying TOPSIS. PCA achieves this by converting the data into a new coordinate system where the first principal component captures the highest variance, and each subsequent component accounts for the highest variance in a direction orthogonal to the previous ones.

The motivation behind using PCA is to eliminate bias in the TOPSIS method by removing correlations among the criteria. By transforming the data into an uncorrelated space, the ideal solutions for TOPSIS are calculated without the influence of correlated dimensions.

By combining PCA's ability to extract uncorrelated contributions with TOPSIS's ranking capabilities, the proposed method effectively eliminates bias present in conventional multi-criteria approaches, which often favor alternatives with strong performance in correlated criteria.

Results and Discussion

The proposed framework was assessed by comparing it with advanced multi-criteria and neural network-based consolidation algorithms. Given that the target system operates in a heterogeneous cloud computing environment, it is essential to account for the diverse computational capacities of cloud hosts. In the simulation, a cloud infrastructure was modeled with a datacenter housing 800 heterogeneous physical machines, including 200 units each of Hitachi HA 8000, DEPO Race X340H, Dell PowerEdge R820, and Dell PowerEdge C6320. An extended version of the CloudSim toolkit and its complete infrastructure were employed as the simulation platform, allowing repeatable experiments on a large-scale virtualized cloud environment with parameters such as response time, cost, and power usage. New modules were developed to simulate PCA and TOPSIS methods.

Dynamic workloads were generated using the real-world open-source Bitbrain dataset, which contains resource consumption metrics from over a thousand VMs hosted on Bitbrain's infrastructure. This data includes logs at 5-minute intervals detailing CPU core requests, CPU usage (MIPS), and RAM characteristics.

The proposed method was compared to state-of-the-art multi-criteria approaches, demonstrating significant improvements, including reductions of 41.5% in energy consumption, 30.13% in SLA violations, 12.9% in response time, 13.4% in running costs, and 58.2% in VM migrations.

Additionally, the performance of the proposed model was compared to neural network-based methods such as Deep Q-Network, Deep Reinforcement Learning, and Asynchronous Advantage Actor-Critic based on Residual Recurrent Neural Network (A3C-R2N2). Results showed that the proposed method achieved the lowest energy consumption, SLA violations, response time, running costs, and VM migrations due to its novel design that enables multi-criteria decision-making and uncorrelated input criteria analysis.

Conclusions

This paper presents an efficient approach to consolidating virtual machines (VMs) in a cloud environment, addressing the challenges posed by diverse VM selection criteria, resource heterogeneity, and the need to account for the uncorrelated effects of these criteria. Previous work overlooks both the variety of decision criteria and the correlations between them, leading to biased choices that favor alternatives with strong performance in two or more related criteria.

Specifically, the paper focuses on the standard four-phase VM consolidation process and introduces a multi-criteria decision-making policy for selecting VMs during the third phase. The proposed method customizes this multi-criteria approach for cloud environments, taking into account various VM selection criteria to ensure more effective decisions by matching resource-intensive VMs with resource-rich cloud hosts.



کارافن

فصلنامه علمی دانشگاه ملی مهارت

بهار ۱۴۰۴، دوره ۲۲، شماره ۱، ۸۱-۱۰۴

آدرس نشریه: <https://karafan.nus.ac.ir/article/229120.html>



<https://doi.org/10.48301/kssa.2025.464305.2935>



رویکرد چند معیاره ادغام ماشین‌های مجازی برای رعایت توافق سطح سرویس و بهره‌وری انرژی در مراکز داده ابری

فرخنده کیائی^{*۱}

^{*۱} - استادیار، گروه مهندسی برق، دانشگاه ملی مهارت، تهران، ایران^۱

اطلاعات مقاله

چکیده

دریافت مقاله : ۱۴۰۳/۰۴/۰۹

بازنگری مقاله : ۱۴۰۳/۰۹/۱۲

پذیرش مقاله : ۱۴۰۳/۱۱/۰۸

کلید واژگان:

رایانش ابری،

روش ترجیح براساس مشابهت به راه حل

ایده‌آل (تاپسیس)،

تحلیل مولفه اصلی،

ادغام،

مدیریت منابع.

ادغام ماشین‌های مجازی سازی یک رویکرد موثر برای دستیابی به بهره‌وری انرژی بهتر در محیط ابری است. با استفاده از مهاجرت زنده ماشین‌های مجازی، چندین ماشین مجازی روی حداقل تعداد از منابع فیزیکی موجود ادغام می‌شوند، به طوری که میزبان‌های باقی‌مانده می‌توانند خاموش شوند. با اینحال ادغام ماشین‌های مجازی نباید منجر به وخامت عملکرد و نقض توافق سطح سرویس شود. در این مطالعه، یک راه حل کارآمد انتخاب ماشین‌های مجازی مناسب برای مهاجرت در ادغام منابع ابری پیشنهاد شده است. روش پیشنهادی از روش تحلیل مولفه اصلی و روش ترجیح براساس مشابهت به راه حل ایده‌آل تاپسیس استفاده می‌کند. این تکنیک‌ها اثرات ناهمبسته معیارهای مختلف را استخراج و امتیازات همه گزینه‌ها را محاسبه می‌کنند. با ترکیب توانایی استخراج مشارکت ناهمبسته معیارها توسط روش تحلیل مولفه اصلی و رتبه‌بندی هوشمند روش تاپسیس، روش پیشنهادی با موفقیت از سوگیری رویکردهای چند معیاره معمول نسبت به انتخاب گزینه‌هایی که ارزیابی‌های خوبی در دو یا چند معیار وابسته دارند، اجتناب می‌کند. نتایج شبیه‌سازی‌ها با استفاده از شبیه‌ساز Cloudsim، کارایی روش پیشنهادی را تایید می‌کند که در مقایسه با روش‌های اخیر تا ۴۱٫۵۴٪، ۳۰٫۱۳٪، ۱۲٫۹۴٪، ۱۳٫۴۲٪ و ۵۸٫۲۲٪ کاهش مصرف انرژی، نقض توافق سطح سرویس، زمان پاسخ، هزینه‌های جاری و تعداد مهاجرت‌های ماشین‌های مجازی را نتیجه می‌دهد.

^۱ نویسنده مسئول: فرخنده کیائی

پست الکترونیکی:

kiaei@shariaty.ac.ir

۱. مقدمه

مدیریت منابع در مراکز داده به دلیل زیربخش های پیچیده و ویژگی های متنوع بارهای کاری ناهمگن بسیار چالش برانگیز است. از این رو، تنظیم دقیق پارامترهای قابل کنترل توسط سامانه های مدیریت منابع به صورت دستی غیرممکن است. علاوه بر این، از آنجا که یک سرور حتی در زمانی که پردازنده آن هیچ کاری انجام نمی دهد، تا ۷۰ درصد از حداکثر توان خود را مصرف می کند، افزایش استفاده از منابع یک پارامتر بهینه سازی حیاتی برای ارائه دهندگان خدمات ابری است. افزایش استفاده از منابع با کاهش تعداد ماشین های فعال، هزینه عملیاتی مصرف انرژی را کاهش می دهد و منابع موجود برای ارائه خدمات به تعداد بیشتری از کاربران استفاده می گردد. در سرویس رایانش ابری زیرساخت به عنوان سرویس^۱، استفاده از منابع معمولاً با ادغام پویا ماشین های مجازی از طریق مهاجرت^۲ آن ها و خاموش کردن ماشین های فیزیکی غیرفعال افزایش می یابد. در این کاربرد، اصطلاح مهاجرت ماشین مجازی به فرآیند انتقال ماشین مجازی در حال اجرا از یک ماشین میزبان فیزیکی به دیگری، بدون ایجاد اختلال در عملکرد ماشین مجازی اشاره دارد.

ادغام پویای ماشین های مجازی یک روش پرکاربرد در مراکز داده ابری برای بهره وری انرژی است. با این حال، ادغام بیش از حد سبب می شود برنامه ها برای منابع محدود در محیط های مشترک موجود رقابت کنند و در نتیجه به طور بالقوه منجر به کاهش کیفیت سرویس و نقض توافق سطح سرویس^۳ برای کاربران می شود. از این رو، باید از راه حل های نوین و کارآمد برای آموختن روابط پیچیده بین پارامترهای مختلف، بهره گرفت و این روش ها را برای حل مشکلات در مدیریت منابع بکار بست.

در این مقاله یک روش جدید برای ادغام ماشین های مجازی در محیط های ابری پیشنهاد می شود. در این راستا، روش تحلیل مولفه اصلی^۴ برای استخراج اثرات ناهمبسته مجموعه معیارهای تصمیم گیری انتخاب ماشین های مجازی بهینه برای مهاجرت استفاده می شود. سپس با استفاده از ماتریس معیارهای ناهمبسته، از روش ترجیح براساس مشابهت به راه حل ایده آل تحت نام روش تاپسیس^۵ برای رتبه بندی و انتخاب بهترین ماشین های مجازی از میان گزینه های موجود استفاده می شود.

مزیت های روش پیشنهادی به طور مختصر در زیر شرح داده شده است:

(۱) برخلاف روش های موجود که تنها معیارهای محدودی را برای ادغام ماشین های مجازی استفاده

می کنند، روش ارائه شده یک سیاست چند معیاره با در نظر گرفتن معیارهای متعددی از جمله مصرف

انرژی، معیار نقض توافق سطح سرویس، زمان پاسخ، هزینه و هم مکانی است.

(۲) روش تاپسیس معمولاً دارای معیارهای وابسته است که ممکن است به تصمیم گیری غیربهینه در

انتخاب ماشین های مجازی برای مهاجرت منجر شود. در روش ارائه شده، این معیارها توسط تحلیل

¹ Infrastructure As A Service (IaaS)

² Migration

³ Service Level Agreement (SLA)

⁴ Principal Component Analysis (PCA)

⁵ Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)

مولفه اصلی به فضای دیگری که وابسته نباشند منتقل می‌شوند تا رتبه‌بندی بهینه‌ای به ماشین‌های مجازی اختصاص داده شود.

(۳) زیربخش‌های نرم افزار CloudSim گسترش داده شده‌اند تا امکان استفاده از پارامترهایی مانند زمان

پاسخگویی و هزینه و همچنین امکان محاسبه تحلیل مولفه اصلی معیارها را فراهم کنند.

در ادامه در بخش ۲ به مرور ادبیات و مقالات مرتبط پرداخته می‌شود. در بخش ۳ روش ارائه شده برای ادغام ماشین‌های مجازی تشریح می‌گردد و در بخش ۴ ارزیابی روش پیشنهادی و مقایسه آن با روش‌های دیگر ارائه می‌شود.

۲. پیشینه پژوهش

حوزه وسیعی از تحقیقات در زمینه مدیریت منابع ابری مرتبط با این پژوهش صورت گرفته است. به طور کلی مسأله مدیریت منابع ابری معمولاً به یک بهینه‌سازی با هدف حداقل کردن مصرف انرژی، هزینه‌ها، زمان پاسخ، نقض توافق سطح سرویس و تعداد مهاجرت‌های ماشین‌های مجازی تبدیل می‌شود [۱].

در [۲] مسأله ادغام ماشین‌های مجازی به چهار بخش تقسیم شده است: (۱) و (۲) تعیین شرایطی که یک میزبان به ترتیب دارای اضافه بار یا کم بار در نظر گرفته می‌شود. (۳) انتخاب ماشین‌های مجازی برای مهاجرت و (۴) پیدا کردن مکان جدید ماشین‌های مجازی انتخاب شده برای مهاجرت. نویسندگان برای مهاجرت ماشین‌های مجازی روش زمان مهاجرت کمینه^۶ را برای انتخاب ماشین‌های مجازی که دارای کمترین زمان مهاجرت هستند، ارائه کردند. رویکرد ادغام پیشنهاد شده در [۳] به زیر مسأله‌های مشابهی پرداخته است و برای مرحله انتخاب ماشین‌های مجازی روش استفاده پیشینه^۷ برای انتخاب ماشین‌های مجازی با بیشترین میزان مصرف CPU ارائه شده است.

در پژوهش انجام گرفته در [۴] با هدف به حداقل رساندن مصرف توان، روش تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس^۸ برای قرار دادن ماشین‌های مجازی ارائه شده است. روش تاپسیس یک الگوریتم تصمیم‌گیری چندمعیاره است که در زمینه‌های گسترده‌ای بکار گرفته شده است [۵؛ ۶].

رویکردی برای ادغام ماشین‌های مجازی با هدف بهبود کیفیت سرویس در [۷] بررسی شده است که در آن بر اساس تاریخچه استفاده از منابع در یک محیط ابری، الگوریتمی کارآمد برای یافتن مکان جدید ماشین‌های مجازی ارائه شده است. در این الگوریتم یک ماشین مجازی در صورتی به میزبان مقصد منتقل می‌شود که استفاده از CPU آن کمترین همبستگی را با استفاده از CPU تمام ماشین‌های مجازی که در آن میزبان قرار دارند، داشته باشد^۹ تا افت عملکرد ناشی از هم‌مکانی ماشین‌های مجازی با منابع موردنیاز مشابه کمینه شود. به طور مشابه در

⁶ Minimum Migration Time (MMT)

⁷ Maximum Usage (MU)

⁸ TOPSIS Power and SLA aware Allocation (TPSA)

⁹ Utilization and Minimum Correlation (UMC)

[۸] نیز دو روش اولین و بهترین برازش^۱ برای قرار دادن ماشین‌های مجازی بر میزبان بکار گرفته شده است که برای اجتناب از هم‌مکانی، همبستگی میان ماشین‌های مجازی بر روی یک میزبان مورد بررسی قرار می‌گیرد. نویسندگان در [۹] نیز مساله ادغام در محیط ابری را در نظر گرفته‌اند و روش کمینه کردن تابع هزینه مشترک متعادل‌سازی بار و مهاجرت^۱ را برای قرار دادن ماشین‌های مجازی طراحی کرده‌اند که تأثیر تأخیر مهاجرت را به حداقل می‌رساند و در عین حال بارهای کاری میزبان‌ها را متعادل می‌کند. در [۱۰] نیز یک الگوریتم زمانبندی اجرای ماشین‌های مجازی مبتنی بر روش برنامه‌ریزی خطی چندهدفه برای اجرای بارهای کاری روی منابع ابری پیشنهاد شده است.

نویسندگان در [۱۱] روشی کارا از نظر انرژی، عملکرد، هزینه، و هم مکانی برای ادغام ماشین‌های مجازی در اجرای چندین بار کاری با خصوصیت هم‌مکانی ارائه کرده‌اند. نویسندگان در [۱۱] روش‌های مختلف قرار دادن و ادغام ماشین‌های مجازی را از نظر سه روش مختلف زمان‌بندی اجرای ماشین‌های مجازی: (۱) متمرکز، (۲) سلسله مراتبی و (۳) توزیع‌شده مورد بحث قرار داده‌اند و در این راستا از یک شبیه‌ساز اختصاصی با نام VMPI aceS بهره گرفته‌اند. روش بهترین برازش کاهشی تعدیل‌شده^۲ در ادغام ماشین‌های مجازی در [۱۲] ارائه شده و محاسبه هزینه‌های مصرف انرژی و مهاجرت مورد مطالعه قرار گرفته است. نویسندگان در [۱۳] یک الگوریتم تخصیص منابع خود تطبیقی^۳ پیشنهاد کرده‌اند که به صورت پویا و کارآمد، منابع را به ماشین‌های مجازی تخصیص می‌دهد. یک روش مدیریت منابع دو مرحله‌ای نیز در [۱۴] ارائه شده است که در آن مرحله اول تعیین اولویت کار و مرحله دوم انتخاب مناسب‌ترین گره میزبان بر اساس زودترین زمان شروع و زودترین زمان پایان است.

در روش ادغام ارائه شده در [۱۵]، روش زمان مهاجرت کمینه برای انتخاب ماشین‌های مجازی برای مهاجرت و روش تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس (با معیارهای تصمیم‌گیری بیشتری نسبت به [۴]) برای قرار دادن ماشین‌های مجازی بکار گرفته شده است. برای جلوگیری از هم‌مکانی ماشین‌های مجازی که برای منابع مشابه روی یک میزبان خاص رقابت می‌کنند، یک ماشین مجازی در صورتی به میزبان مقصد منتقل می‌شود که استفاده از CPU توسط آن کمترین همبستگی را با استفاده از CPU در تمام ماشین‌های مجازی که در آن میزبان قرار دارند، داشته باشد. یکی دیگر از رویکردهای ادغام ماشین‌های مجازی در [۱۶] توسعه داده شده است و روشی به نام روش ادغام مبتنی بر انرژی و عملکرد^۴ برای انتخاب ماشین‌های مجازی برای مهاجرت ارائه شده است.

با اینحال، الگوریتم‌های ادغام چند معیاره ارائه شده همگی تعداد محدودی از معیارها را در نظر می‌گیرند و معیارهای تصمیم‌گیری بکار گرفته توسط آنها همبسته هستند. این روش‌ها فاقد بخشی برای استخراج اثرات ناهمبسته معیارها هستند که انجام پژوهش موردنظر در این مقاله را ترغیب می‌کند.

¹ First fit/Best Fit

¹ Load Balancing Joint Migration Cost[‡] (LBJC)

¹ Modified Best Fit Decreasing (MBF[§])

¹ Self-Adaptive Resource Allocation (§ARA)

¹ Energy Power Consolidation (EPC) ⁴

روش‌های مبتنی بر شبکه‌های عصبی نیز برای مدیریت منابع ابری بکار گرفته شده‌اند. در روش‌های مانند یادگیری Q یک تابع مقدار (ارزش) Q برای هر حالت از محیط ابری با استفاده از شبکه‌های عصبی عمیق^۵ تخمین زده می‌شود [۱۷]. این ارزش در طول فرآیند آموزش بروزرسانی می‌شود و در واقع نشان دهنده پاداش تجمعی مورد انتظار است. روش‌های یادگیری تقویتی مبتنی بر شبکه‌های عصبی عمیق^۶ نیز برای مدیریت منابع ابری بکار گرفته شده‌اند [۱۸]. در [۱۹] یک ساختار شبکه عصبی بازگشتی مانده که در آن از روش عملگر-منتقد غیرسنکرون^۷ برای آموزش استفاده شده است برای مدیریت منابع ابری ارائه شده است. با اینحال ساختار شبکه‌های عصبی ارائه شده دارای تعداد گره‌های لایه آخر ثابت است که موجب می‌شود تنها تعداد ثابتی از میزبان‌ها در خروجی شبکه در نظر گرفته شود و در نتیجه مقیاس پذیر نیستند.

۳. رویکرد ادغام ماشین‌های مجازی در محیط ابری

روش ارائه شده در این پژوهش مساله ادغام ماشین‌های مجازی را به چهار بخش تقسیم کرده است:

(۱) تعیین زمانی که یک میزبان دارای اضافه بار در نظر گرفته می‌شود: تعیین میزبان‌های با اضافه بار با روش رگرسیون محلی^۸ انجام می‌شود [۲].

(۲) تعیین زمانی که یک میزبان کم بار در نظر گرفته می‌شود: تعیین میزبان‌های کم بار با روش (آستانه گیری پویای) ساده‌ای^۹ صورت می‌گیرد [۲] که در آن مقدار آستانه تابعی از میزان مصرف CPU است. بدین ترتیب مقدار آستانه در بازه‌هایی که بارکاری کلی سیستم کمتر است، پایین آورده می‌شود تا میزبان‌های کم بار با دقت بیشتری شناسایی شوند.

(۳) انتخاب ماشین‌های مجازی برای مهاجرت: روشی چندمعیاره با قابلیت یادگیری اثرات ناهمبسته معیارها برای این مرحله بکار گرفته می‌شود که در بخش ۳،۲ ارائه می‌گردد.

(۴) پیدا کردن میزبان‌های جدید ماشین‌های مجازی انتخاب شده برای مهاجرت: تعیین میزبان‌های جدید برای قرار دادن ماشین‌های مجازی با روش تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس صورت می‌گیرد [۱۵].

در ادامه در ابتدا معماری سامانه هدف تعریف می‌شود و سپس روش پیشنهادی برای انتخاب ماشین‌های مجازی برای مهاجرت ارائه و مورد ارزیابی قرار داده می‌شود.

۱،۳. معماری سامانه هدف

مدل سامانه هدف شامل سه بخش مهم است: (۱) زیرساخت ابری، (۲) کاربران با بارهای کاری مختلف و (۳) بخش ادغام پیشنهادی. یک نمای کلی از مدل سامانه در شکل ۱ نشان داده شده است. بخش ادغام خودش شامل

^۱ Deep Q Network (DQN) 5

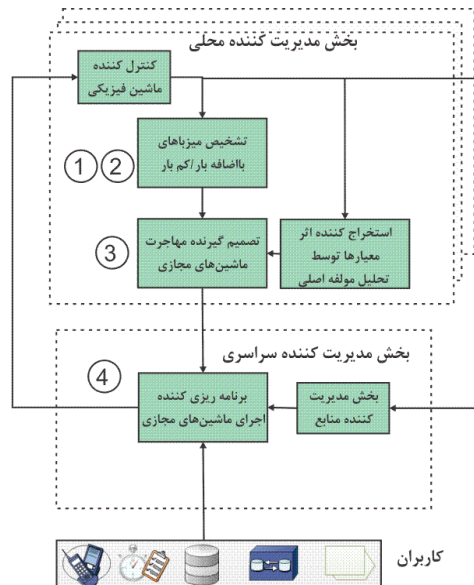
^۱ Deep Reinforcement Learning (DRL)^۹

^۱ Asynchronous Advantage Actor-Critic Residual Recurrent Neural Network (A3C-R2N2)

^۱ Local Regression (LR) 8

^۱ Simple Method (SM) 9

"بخش مدیریت کننده محلی" و "بخش مدیریت کننده سراسری" است که اولی روی همه گره های میزبان (ماشین های فیزیکی) اجرا می شود و دومی روی یک گره میزبان مرکزی (ماشین فیزیکی مرکزی) اجرا می شود. اعداد ۱ تا ۴ نشان داده شده در دایره ها در شکل ۱ با چهار مرحله رویکرد ادغام ماشین های مجازی اشاره شده، مطابقت دارند.



شکل ۱. معماری سامانه پیشنهادی.

در معماری سامانه هدف، زیرساخت ابری از گره های ابری با منابع ناهمگن تشکیل شده است. این گره ها (ماشین های فیزیکی) با پارامترهای بسیار متفاوتی از جمله توان محاسباتی مختلف برحسب میلیون ها دستورالعمل در ثانیه فعالیت می کنند و میزبان کاربران با برنامه های مختلف هستند که چندین ماشین مجازی را اجرا می کنند. این فرآیند منجر به بار کاری ترکیبی پویا بر هر میزبان می شود. بدین ترتیب ارائه دهندگان خدمات ابری می توانند از روش پیشنهادی براساس معیارهای مختلف برای شناسایی ماشین های مجازی مناسب برای مهاجرت در یک محیط ابری استفاده کنند و با ادغام آنها، تعداد میزبان های فعال را کاهش دهند. بدین ترتیب هنگام برنامه ریزی اجرای ماشین مجازی، می توان با استفاده از روش پیشنهادی، پارامترهای مختلف مانند مصرف انرژی و توافق سطح سرویس را بر اساس نیازهای کاربران و ارائه دهندگان خدمات ابری بهینه کرد.

بخش مدیریت کننده محلی در هر گره میزبان فعال وجود دارد و از چهار جزء اصلی تشکیل شده است که به شرح زیر معرفی می شوند:

(۱) کنترل‌کننده ماشین فیزیکی: برنامه‌ای است که در هر ماشین فیزیکی وجود دارد و اطلاعاتی را در مورد میزان استفاده از منابع، ظرفیت منابع، تعداد ماشین‌های مجازی و همچنین انواع ماشین‌های مجازی در یک ماشین فیزیکی جمع‌آوری می‌کند.

(۲) تشخیص میزان‌های کم بار/با اضافه بار: این بخش میزان‌های کم بار/با اضافه بار را با استفاده از روش آستانه‌گیری دوگانه که از مقادیر آستانه پایین (۲۰٪) و بالا (۸۰٪) بهره می‌گیرد، تعیین می‌کند. این اطلاعات به بخش تصمیم‌گیرنده مهاجرت ماشین‌های مجازی ارسال می‌شود.

(۳) بخش استخراج‌کننده اثر معیارها با تحلیل مولفه اصلی: در این بخش تحلیل مولفه اصلی برای استخراج اثرات ناهمبسته معیارهای موجود در ماتریس تصمیم‌گیری اعمال می‌شود. خروجی روش تحلیل مولفه اصلی در فضای ناهمبسته به عنوان معیار ورودی روش تاپسیس در روش پیشنهادی (ارائه شده در بخش بعد) برای انتخاب ماشین‌های مجازی برای مهاجرت استفاده می‌شود. معیارهای ورودی روش تحلیل مولفه اصلی، معیارهای فهرست شده در جدول ۱ هستند که توسط کنترل‌کننده ماشین فیزیکی جمع‌آوری می‌شود. روش تحلیل مولفه اصلی این معیارهای ورودی را به فضای ناهمبسته خروجی نگاشت می‌کند.

(۴) تصمیم‌گیرنده مهاجرت ماشین‌های مجازی: این بخش ماشین‌های مجازی مناسب برای مهاجرت را با استفاده از روش پیشنهادی (ارائه شده در بخش بعد) تعیین می‌کند. هدف از مهاجرت ماشین‌های مجازی، بهبود استفاده از منابع با توزیع یکنواخت ماشین‌های مجازی در سراسر میزان‌ها و به حداقل رساندن تعداد ماشین‌های فیزیکی فعال است. ماشین‌های مجازی مناسب برای مهاجرت از میزان‌های با اضافه بار انتخاب می‌شوند تا مصرف انرژی میزان تعدیل شود. علاوه بر این، ماشین‌های مجازی را می‌توان از یک میزان با بار کم منتقل کرد تا بتوان آن را خاموش کرد تا تعداد میزان‌ها به حداقل برسد. هنگامی که ماشین‌های مجازی قابل مهاجرت یافت می‌شوند، این بخش تصمیمات مهاجرت را به بخش برنامه‌ریزی‌کننده اجرای ماشین مجازی اطلاع می‌دهد.

بخش مدیریت‌کننده سراسری بر یک گره میزان جداگانه تنظیم می‌شود و مقصدهای مناسبی را برای ماشین‌های مجازی انتخاب شده برای مهاجرت تعیین می‌کند. این بخش از اجزای زیر تشکیل شده است:

(۱) بخش مدیریت‌کننده منابع: این بخش اطلاعات مربوط به ماشین‌های فیزیکی را جمع‌آوری می‌کند و معیارهای تصمیم‌گیری مانند مصرف توان و معیار نقض توافق سطح سرویس را به بخش برنامه‌ریزی‌کننده اجرای ماشین‌های مجازی که وظیفه‌اش پیدا کردن میزان مناسب برای ماشین‌های مجازی است، ارسال می‌کند.

(۲) بخش برنامه‌ریزی‌کننده اجرای ماشین‌های مجازی: این بخش مکان ماشین‌های مجازی را بر روی میزان‌ها تعیین می‌کند. تصمیمات این بخش بر اساس اطلاعات موجود از بخش مدیریت‌کننده منابع است و از روش تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس [۱۵] برای یافتن مکان مناسب (مقصد مهاجرت) برای هر ماشین مجازی بر یک ماشین فیزیکی استفاده می‌کند. هنگامی که این بخش مکانی را پیدا می‌کند، به کنترل‌کننده ماشین فیزیکی اطلاع می‌دهد تا ماشین

مجازی را راه اندازی کند و بروزرسانی اطلاعات منابع موجود را انجام دهد. شکل ۲ نمودار دنباله‌ای^{۲۱} برای معماری پیشنهادی را نشان می‌دهد.



شکل ۲. نمایش نمودار دنباله‌ای برای روش ادغام پیشنهادی

همانطور که پیش‌تر گفته شد، روش مدیریت منابع ابری پیشنهادی روشی است که براساس اهداف از پیش تعیین شده، ماشین‌های مجازی مناسب برای مهاجرت را تعیین می‌کند. جزئیات روش پیشنهادی در بخش بعد شرح داده می‌شود.

۲.۳. روش ارائه شده انتخاب ماشین‌های مجازی برای مهاجرت با استخراج اثرات ناهمبسته معیارها

روش پیشنهادی از روش تحلیل مولفه اصلی و همچنین روش تاپسیس به عنوان یک الگوریتم چند معیاره استفاده می‌کند و چهار معیار نشان داده شده در جدول ۱ را در تصمیم‌گیری خود در نظر می‌گیرد.

جدول ۱. معیارهای مورد استفاده در انتخاب ماشین‌های مجازی برای مهاجرت در محیط ابری

ردیف	نام اختصاری	پارامتر	توصیف	هزینه/سود
۱	VMM	ظرفیت محاسباتی در ماشین مجازی	ظرفیت محاسباتی ماشین مجازی برحسب MIPS	هزینه

² sequence diagram

سود	مصرف CPU ماشین مجازی	استفاده منابع محاسباتی در ماشین مجازی	VMCU	۲
سود	همبستگی منابع یک ماشین مجازی با دیگر ماشین های مجازی بر یک ماشین فیزیکی	همبستگی منابع	RC	۴
هزینه	ظرفیت ذخیره سازی یک ماشین مجازی	حافظه ماشین مجازی	VMS	۵

این روش امتیازات همه ماشین های مجازی را که برای مهاجرت مناسب هستند با استفاده از روش تاپسیس محاسبه می کند و ماشین های مجازی که بالاترین امتیاز را دارند انتخاب می کند. معیارهای در نظر گرفته شده در این روش همانطور که در جدول ۱ نشان داده شده است، از نوع هزینه یا سود هستند. روش پیشنهادی امتیاز ماشین های مجازی را به گونه ای محاسبه می کند تا شرایط زیر در پاسخ وجود داشته باشد:

(۱) ظرفیت محاسباتی ماشین مجازی انتخاب شده باید حداقل باشد: انتخاب ماشین مجازی با ظرفیت محاسباتی بالا برای مهاجرت، خطر کمبود منابع برای ماشین های مجازی موجود بر میزبان را افزایش می دهد.

(۲) میزان استفاده منابع در ماشین مجازی انتخاب شده باید حداکثر باشد: انتخاب چنین ماشین مجازی با اولویت بالاتر برای مهاجرت، احتمال کمبود منابع و بر این اساس، احتمال نقض توافق سطح سرویس را کاهش می دهد.

(۳) همبستگی منابع بین ماشین مجازی انتخاب شده برای مهاجرت و سایر ماشین های مجازی در همان میزبان باید حداکثر باشد: با افزایش همبستگی بین برنامه هایی که از منابع یکسان در میزبان استفاده می کنند، احتمال اضافه بار در میزبان را افزایش می دهد. برای محاسبه ضرایب همبستگی، برای هر جفت ماشین مجازی (ماشین مجازی مورد نظر و هر ماشین مجازی دیگر روی میزبان مشابه)، ضریب همبستگی پیرسون برای استفاده از منابع ماشین های مجازی محاسبه می شود. سپس میانگین همبستگی ماشین مجازی هدف با تمام ماشین های مجازی دیگر روی میزبان به عنوان یک معیار تصمیم گیری در نظر گرفته می شود.

(۴) ظرفیت ذخیره سازی ماشین مجازی انتخاب شده باید حداقل باشد: با انتخاب یک ماشین مجازی با کمترین ظرفیت ذخیره سازی، سربار مهاجرت به دلیل انتقال مقادیر ذخیره سازی شده کاهش می یابد.

معیارهای ذکر شده در بالا، برای ساخت ماتریس تصمیم گیری چند معیاره مطابق زیر بکار گرفته می شوند:

$$X = \begin{bmatrix} RC^1 & VMCU^1 & VMM^1 & VMS^1 \\ RC^2 & VMCU^2 & VMM^2 & VMS^2 \\ \vdots & \vdots & \vdots & \vdots \\ RC^n & VMCU^n & VMM^n & VMS^n \end{bmatrix}, \quad (1)$$

که در آن n تعداد گره های میزبان است. قبل از استفاده از X به عنوان ورودی روش تاپسیس، این معیارها به ورودی روش تحلیل مولفه اصلی ارسال می شوند و الگوریتم تحلیل مولفه اصلی برای استخراج اثرات ناهمبسته معیارهای موجود در ماتریس تصمیم گیری اعمال می شود [۲۰]. باید به این نکته توجه داشت که با داشتن یک ماتریس معیارهای ورودی با ابعاد $d \times n$ ($d = ۴$)، برخی از معیارها احتمالاً دارای همبستگی هستند، زیرا آنها از یک ماشین مجازی گرفته شده اند. روش تحلیل مولفه اصلی از یک تبدیل متعامد برای نگاشت به یک دستگاه مختصات جدید و بدست آوردن تعداد نسبتاً کمی معیارهای ناهمبسته که مولفه های اصلی هستند، استفاده می کند. اولین مولفه اصلی به جهتی با بیشترین واریانس اشاره می کند و مولفه های اصلی بعدی هر کدام به بزرگترین جهت واریانس اشاره می کنند که متعامد به همه مولفه های قبلی هستند.

استفاده از تحلیل مولفه اصلی در کاربرد مدیریت خودکار ابر به طور گسترده ای در زمینه تشخیص ناهنجاری (به عنوان مثال در [۲۱])، مورد مطالعه قرار گرفته است. با اینحال، هیچ مطالعه و مقاله قبلی در مورد بکارگیری تحلیل مولفه اصلی برای پیاده سازی ادغام در محیط های ابری وجود ندارد.

در روش تحلیل مولفه اصلی، تجزیه مقادیر ویژه بر ماتریس کوواریانس انجام می شود و مجموعه ای از بردارهای ویژه $V = (v_1, v_2, \dots, v_d)$ به دست می آید که بر اساس مقادیر ویژه آنها مرتب شده اند. این بردارهای ویژه محورهای جدیدی را در دستگاه مختصات جدید تشکیل می دهند که اولین محور اصلی v_1 به جهتی اشاره دارد که بیشترین واریانس را دارد و محورهای اصلی بعدی هر کدام به بزرگترین جهت واریانس متعامد نسبت به قبلی اشاره می کنند. مقادیر ویژه مربوطه $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$ هستند.

زیرفضای اصلی به عنوان فضایی که توسط اولین k محور اصلی در V یعنی $P_1 = (v_1, v_2, \dots, v_k)$ دربرگرفته شده است، تعریف می شود و فضایی که توسط بقیه محورهای اصلی پوشانده شده است به عنوان فضای مانده شناخته می شود. روش درصد واریانس تجمعی $\sum_{i=1}^k \lambda_i / \sum_{i=1}^d \lambda_i$ و عملکرد خوب آن برای تعیین k استفاده می شود. برای اولین مؤلفه های اصلی، $CPV(j) = 100\% (\sum_{i=1}^j \lambda_i) / (\sum_{i=1}^d \lambda_i)$ را داریم و k براساس رابطه $k = \arg \min_j (CPV(j) > 90\%)$ انتخاب می شود [۲۰].

به طور خلاصه، روش پیشنهادی شامل سه مرحله است:

(۱) ساخت ماتریس تصمیم گیری چند معیاره X از تمام معیارهای اختصاص داده شده به گزینه ها در هر

بازه زمانی، همانطور که در معادله (۱) نشان داده شده است.

(۲) ساخت ماتریس معیارهای فضای ناهمبسته Z با انجام تحلیل مولفه اصلی بر X و تعیین اولین k مؤلفه

اصلی برای ایجاد یک ماتریس وزن داده شده به صورت زیر

$$Z = XP_1, \quad (2)$$

که در آن وزن ها بردارهای ویژه مرتبط با مؤلفه های اصلی $P_1 = (v_1, v_2, \dots, v_k)$ هستند.

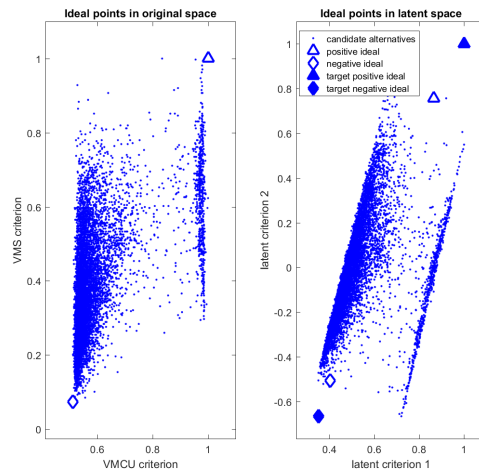
(۱) انجام تاپسیس برای کسب امتیاز برای گزینه نام توسط معادله زیر،

$$S = \frac{\sqrt{\sum_{j=1}^k (z_j^i - z_j^-)^2}}{\sqrt{\sum_{j=1}^k (z_j^i - z_j^+)^2} + \sqrt{\sum_{j=1}^k (z_j^i - z_j^-)^2}} \quad (3)$$

که در آن z_j^i درایه ماتریس فضای ناهمبسته Z پس از نرمال سازی (با تقسیم هر ورودی بر حداکثر مقدار هر ستون) است. علاوه بر این، z_j^+ راه حل مثبت ایده آل (حداکثر z_j^i) و z_j^- راه حل منفی ایده آل (حداقل z_j^i) است $(i = 1, 2, \dots, n \text{ و } j = 1, 2, \dots, k)$.

ایده اصلی استفاده از تحلیل مولفه اصلی به عنوان استخراج کننده اثرات ناهمبسته معیارهای ورودی در روش تاپسیس آن است که نقاط ایده آل مثبت و منفی هد با احتمال زیادی با کمی سوگیری و خطا به دلیل همبستگی میان ابعاد، در فضای اصلی محاسبه می شوند. بنابراین با تبدیل ماتریس معیارها به فضای پنهان ناهمبسته جدید، مختصات اصلی به گونه ای چرخانده می شوند که ایده آل های مثبت و منفی جدید راه حل های ایده آل در این فضای ناهمبسته باشند.

نمونه ای از حالت دوبعدی با $d = 2$ شامل میزان استفاده از منابع محاسباتی (معیار VMCU در جدول ۱) و حافظه ذخیره سازی (معیار VMs در جدول ۱) برای یک بارکاری تصادفی در مجموعه داده [۲۲] Bi t brain (که در این مطالعه بررسی شده) در شکل ۳ نشان داده شده است. همانطور که در شکل ۳ (بخش سمت چپ) مشخص است، این دو معیار همبستگی بالایی از خود نشان می دهند. اگر تاپسیس را بر داده های اصلی بکار ببریم، ایده آل های مثبت و منفی مطابق آنچه در نمودار سمت چپ نشان داده شده است، می باشند. در مقایسه، اگر این روش را در فضای جدید ناهمبسته بکار ببریم، هر دو ایده آل مثبت و منفی هدف در نمودار سمت راست نشان داده شده اند. به راحتی در نمودار سمت راست می توان یافت که تاپسیس اعمال شده بر روی داده های اصلی منجر به ایده آل های مثبت و منفی مطلوب نمی شود چرا که نگاشت این نقاط به فضای جدید با ایده آل های مثبت و منفی هدف در آن فضا مطابقت ندارد. در نتیجه، در ادامه محاسبات در روش تاپسیس، معیار فاصله به درستی محاسبه نمی شود، که منجر به رتبه بندی گزینه های متفاوت با هدف مطلوب نمی گردد. به عبارت دیگر، فضای همبسته اصلی، خطاهایی را در محاسبه ایده آل های مثبت و منفی وارد می کند.



شکل ۲. نمایش گرافیکی ایده تبدیل فضای روش تحلیل مولفه اصلی بر معیارهای تاپسیس در فضای دو بعدی.

۴. ارزیابی

در این بخش، ارزیابی روش ارائه شده در این مقاله را مورد بحث قرار می دهیم. مشخصات مصرف توان میزبان های مورد بررسی و همچنین پیکربندی میزبان ها در جدول ۲ نشان داده شده است. میزبان ها در دو ردیف اول با منابع محدود (Intel i3 و Intel i5) و میزبان ها در ردیف سوم و چهارم جدول با منابع قدرتمند (Xeon) به صورت ناهمگون لحاظ شده اند. در این بخش تعداد ۶۰۰ میزبان در پیاده سازی مورد بررسی قرار گرفته شده است. زمان پاسخ برای میزبان های دو ردیف اول ۱ میلی ثانیه و برای میزبان ها در ردیف سوم و چهارم ۱۰ میلی ثانیه در نظر گرفته شده است [۲۳].

جدول ۲. مشخصات میزبان ها به همراه اطلاعات مصرف توان آنها

مشخصات مصرف توان (Watts)						پهنای دیسک (MB/s)	هزینه (\$/hr)	پهنای باند (GB/s)	RAM (GB)	فرکانس (Hz)	هسته	مدل CPU	میزبان
100%	80%	60%	40%	20%	0%								
63.2	55.7	45.6	39.6	33.7	24.3	76	0.11	0.1	8	1800	2	Intel i3 3.0 GHz	Hitachi HA 8000
131	124	117	107	94.3	83.2	49	0.23	1	16	2000	4	Intel i5 3.2 GHz	DEPO Race X340H
446	358	268	218	167	110	49	3.47	1	48	2000	32	Intel Xeon	Dell PowerEdge

												2.6 GHz	R820
1229	924	705	589	449	210	1024	6.94	1.5	64	2660	64	Intel Xeon 2.3 GHz	Dell PowerEdge C6320

علاوه براین، روش‌های مورد بررسی بر یک مجموعه داده و دنباله حجم کار واقعی پیاده‌سازی شده‌اند. در این راستا از مجموعه داده واقعی Bitbrain استفاده شده است [۲۲]. این داده شامل معیارهای استفاده از منابع ماشین‌های مجازی است که در یک زیرساخت مرکز داده واقعی میزبانی می‌شوند و شامل ثبت اطلاعات بیش از هزار ماشین مجازی بر انواع مختلف میزبان است. به طور مشخص، این داده شامل گزارش‌های اطلاعات بار کاری در فواصل ۵ دقیقه‌ای از جمله تعداد هسته‌های پردازشی درخواستی، میزان استفاده از منابع محاسباتی و ویژگی‌های حافظه مورد استفاده است.

۴.۱. مقایسه روش پیشنهادی با روش‌های چندمعیاره

در این بخش، روش پیشنهادی با پنج سناریوی متفاوت از روش‌های چندمعیاره دیگر پیشنهاد شده در برخی مقالات و تعمیم آنها مقایسه و ارزیابی می‌شود (جدول ۳). سناریوها از یک تا شش شماره گذاری شده‌اند، و راه حل نهایی پیشنهادی سناریو ۶ است. در چهار ستون آخر جدول ۳، روش‌های بکارگرفته شده در چهار بخش رویکرد ادغام در هریک از سناریوهای ارزیابی شده نشان داده شده است. شش سناریوی در نظر گرفته شده به شرح زیر تعریف می‌شوند:

جدول ۳. مشخصات سناریوهای بکارگیری روش‌های مختلف در ارزیابی

ردیف	رویکردها	تشخیص میزبان با اضافه بار	تشخیص میزبان کم بار	انتخاب ماشین مجازی	تعیین میزبان جدید
۱	[۱۶]	آستانه‌گیری ایستا	آستانه‌گیری ایستا	ادغام مبتنی بر انرژی و عملکرد	بهترین برآزش کاهش‌ی تعدیل‌شده
۲	[۹]	آستانه‌گیری ایستا	-	استفاده بیشینه	کمینه کردن تابع هزینه مشترک متعادل سازی بار و مهاجرت
۳	[۱۵]	رگرسیون محلی	روش (آستانه‌گیری پویا) ساده	زمان مهاجرت کمینه	تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس

۴	گسترش [۱۵] با روش انتخاب ماشین مجازی پیشنهادی	آستانه گیری ایستا	آستانه گیری ایستا	روش پیشنهادی	بهترین برازش کاهشی تعدیل شده
۵	گسترش [۸] با روش انتخاب ماشین مجازی پیشنهادی	آستانه گیری ایستا	-	روش پیشنهادی	کمینه کردن تابع هزینه مشترک متعادل سازی بار و مهاجرت
۶	گسترش [۱۴] روش انتخاب ماشین مجازی پیشنهادی	رگرسیون محلی	روش (آستانه گیری پویا) ساده	روش پیشنهادی	تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس

سناریوی ۱ راه حل ارائه شده در [۱۶] را نشان می‌دهد که در آن روش آستانه‌گیری ایستا با آستانه ۷۰ درصد برای تعیین میزان کم بار و آستانه ۸۰ درصد برای تعیین میزان‌های بااضافه بار، روش ادغام مبتنی بر انرژی و عملکرد برای انتخاب ماشین‌های مجازی برای مهاجرت و روش بهترین برازش کاهشی تعدیل شده برای قرار دادن ماشین‌های مجازی استفاده شده است. روش ادغام مبتنی بر انرژی و عملکرد از حاصلضرب معیار مصرف انرژی و معیار عملکرد (معکوس زمان اجرا) به عنوان معیار رتبه بندی برای اولویت بندی آن دسته از ماشین‌های مجازی استفاده می‌کند که مصرف انرژی و عملکرد را بیشتر ارتقا می‌دهند.

سناریوی ۲ رویکرد تصمیم‌گیری مهاجرت ارائه شده در [۹] را نشان می‌دهد که شامل روش آستانه‌گیری ایستا برای هر دو معیار میزان استفاده از منابع و تنوع بار برای تعیین میزان‌ها با اضافه بار، روش استفاده بیشینه برای انتخاب ماشین‌های مجازی که باید مهاجرت کنند، روش کمینه کردن تابع هزینه مشترک متعادل سازی بار و مهاجرت برای قرار دادن ماشین‌های مجازی است که تأثیر تأخیر مهاجرت را به حداقل می‌رساند و در عین حال بار میزان را متعادل می‌کند.

سناریوی ۳ رویکرد ارائه شد در [۱۵] است که شامل رگرسیون محلی برای تعیین میزان‌های بااضافه بار، روش (آستانه‌گیری پویا) ساده برای تعیین میزان های کم بار، روش زمان مهاجرت کمینه برای شناسایی ماشین‌های

مجازی برای مهاجرت، و روش تخصیص منابع آگاه از توان و توافق سطح سرویس مبتنی بر روش تاپسیس برای یافتن مکان‌های جدید برای ماشین‌های مجازی است.

سناریوی ۴، ۵ و ۶ به ترتیب تعمیمی از سناریوی ۱، ۲ و ۳ هستند که در آن مرحله انتخاب ماشین‌های مجازی برای مهاجرت توسط روش پیشنهادی با در نظر گرفتن معیارهای متعدد در محیط ابری و همچنین یادگیری اثرات ناهمبسته معیارها توسط تحلیل مولفه اصلی گسترش داده شده است.

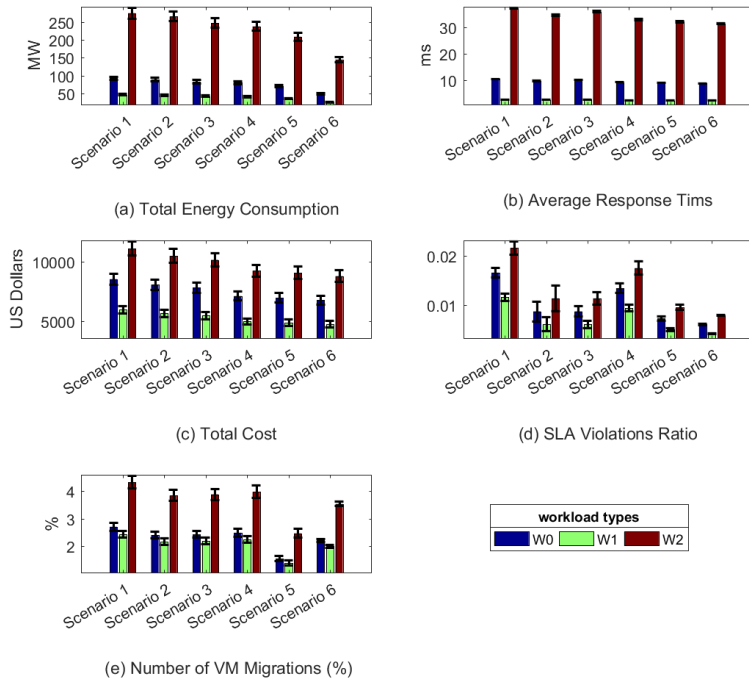
ده آزمایش به طور جداگانه برای ده روز انجام شده است و نتایج شش رویکرد ذکر شده برای معیارهای مصرف انرژی، نقض توافق سطح سرویس، زمان پاسخ، هزینه و تعداد مهاجرت ماشین‌های مجازی در جدول ۴ گزارش شده است.

جدول ۴. نتایج ارزیابی روش‌های مختلف

شماره سناریو	مصرف انرژی (kW)	زمان پاسخ (ms)	هزینه (US Dollars)	معیار نقض توافق سطح سرویس	تعداد مهاجرت ماشین‌های مجازی (%)
	۹۱۷,۹۵	۱۰,۵	۸۵۵۰	۰,۰۱۶۵	۲,۷
۲	۸۸۸,۳۹	۹,۸	۸۰۸۸	۰,۰۰۸۷	۲,۴
۳	۸۲۶,۳۹	۱۰,۲	۷۸۲۷	۰,۰۰۸۷	۲,۴۲
۴	۷۹۳,۸۷	۹,۳۶	۷۱۲۸	۰,۰۱۳۴	۲,۴۸
۵	۶۹۷,۷۲	۹,۱۲	۶۹۸۴	۰,۰۰۷۳	۱,۸۳
۶	۴۸۳,۰۸	۸,۸۸	۶۷۷۶	۰,۰۰۶۱	۱,۹

با توجه به جدول، می‌توان استنباط کرد که راه حل پیشنهادی منجر به عملکرد بهتری در مورد مصرف انرژی، معیار نقض توافق سطح سرویس، زمان پاسخ و تعداد مهاجرت ماشین‌های مجازی در مقایسه با سایر سناریوها می‌شود. به طور دقیق‌تر، می‌توان از جدول ۴ استنباط کرد که پذیرش روش پیشنهادی (سناریوی ۴) به ترتیب، منجر به ۱۳,۵۱٪، ۱۸,۸۴٪، ۱۰,۳۴٪، ۱۶,۶۳٪ و ۷,۹۵٪ کاهش مصرف انرژی، نقض توافق سطح سرویس، زمان پاسخ، هزینه، تعداد مهاجرت‌های ماشین‌های مجازی در مقایسه با سناریوی ۱ می‌شود. همچنین، از جدول ۴ می‌توان استنباط کرد که پذیرش روش پیشنهادی (سناریوی ۵) به ترتیب، منجر به ۲۱,۴۶٪، ۳۹٪، ۷,۳۱٪، ۱۳,۶۴٪ و ۳۵,۸۹٪ کاهش مصرف انرژی، نقض توافق سطح سرویس، زمان پاسخ، هزینه و تعداد مهاجرت در مقایسه با سناریوی ۲ می‌شود. در مقایسه با سناریوی ۳ نیز می‌توان دید که پذیرش روش پیشنهادی (سناریوی ۶)

به ترتیب، منجر به ۴۱٫۵۴٪، ۳۰٫۱۳٪، ۱۲٫۹۴٪، ۱۳٫۴۲٪ و ۵۸٫۲۲٪ کاهش مصرف انرژی، نقض توافق سطح سرویس، زمان پاسخ، هزینه و تعداد مهاجرت می شود.



شکل ۳. مقایسه روش پیشنهادی با سایر رویکردها برای انواع بارهای کاری مختلف.

سپس برای بررسی عملکرد روش پیشنهادی بر انواع بارهای کاری مختلف از مجموعه داده های Google شامل زمان های اجرا و اولویت های کار است، استفاده می شود. نتایج تمام آزمایش ها بر داده ترکیبی (مجموعه داده Bi tran + Google) و مربوط به سه نوع بار کاری W، W1 و W2 که به وظایف با سه اولویت مختلف (۰، ۴، ۹) تعلق دارند، به صورت گرافیکی در شکل ۳ نشان داده می شوند.

از شکل ۳، مشاهده می شود که برای همه بارهای کاری مانیتورینگ/کنسول (W)، بارهای کاری کوتاه مدت (W1) و بارهای کاری طولانی مدت (W2) روش پیشنهادی عملکرد برتری را در مقایسه با سایر رویکردها نشان می دهد. بارهای کاری W با زمان اجرای سریع و چرخه های شروع/توقف مکرر شناخته می شوند. رویکرد ادغام پیشنهادی، مصرف انرژی را با بهینه سازی قرار دادن این بارهای کاری و به حداقل رساندن تعداد میزبان های فعال کاهش می دهد. اگرچه، پیش بینی نادرست زمان اجرا در رویکرد ۵ به طور بالقوه تعداد کل مهاجرت های آن را افزایش می دهد، رویکرد بهینه پیشنهادی می تواند این اثر را کاهش دهد تا از افزایش تعداد مهاجرت ها جلوگیری کند و در عین حال مصرف انرژی را به حداقل برساند. بارهای کاری W2 با زمان اجرای طولانی مستلزم عملکرد پایدار در

دوره‌های طولانی هستند. ادغام برای این بارهای کاری مستلزم برقراری تعادل میان کارایی انرژی و تضمین عملکرد است. رویکرد پیشنهادی با موفقیت این بارهای کاری را به گونه‌ای ادغام می‌کند که از کاهش عملکرد بدون افزایش دادن تعداد مهاجرت‌ها در مقایسه با سایر رویکردها جلوگیری می‌کند. نتایج به‌دست‌آمده، کارایی سناریوی پیشنهادی را برای ادغام در محیط ابری تأیید می‌کند. این مشاهدات را می‌توان با این واقعیت توصیف کرد که راه‌حل پیشنهادی با استفاده از تحلیل مولفه اصلی اثرات ناهمبسته معیارهای متعددی را لحاظ می‌کند و به طور قابل توجهی نتایج را بهبود می‌بخشد.

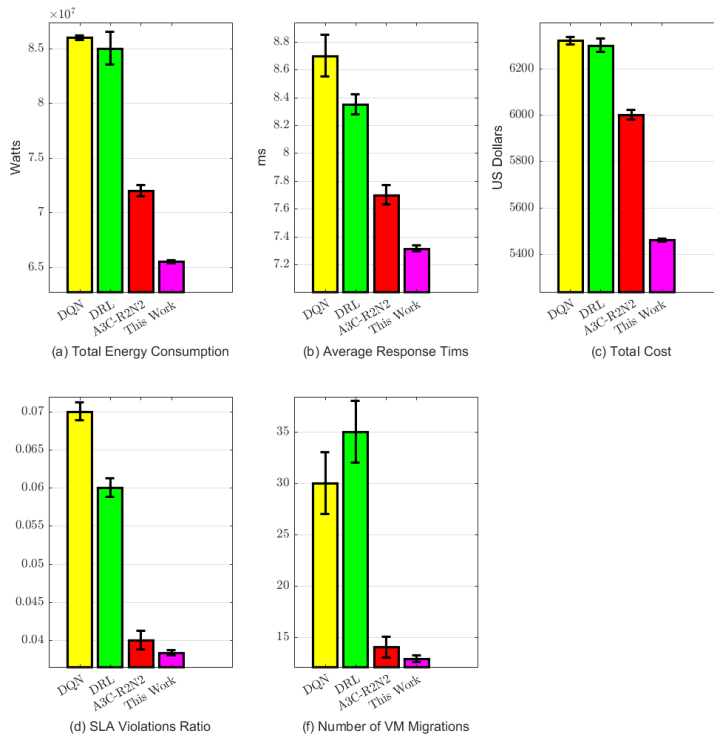
۴,۲. مقایسه روش پیشنهادی با روش‌های مبتنی بر شبکه‌های عصبی

مقایسه‌ای از نتایج روش پیشنهادی و روش‌هایی که از ساختار شبکه‌های عصبی بهره می‌برند برای معیارهای مذکور در شکل ۴ نشان داده شده است. روش‌های مورد مقایسه به شرح زیر هستند:

- روش شبکه عصبی عمیق Q یا DQN: رویکرد استاندارد یادگیری Q که از شبکه عصبی عمیق در پیاده‌سازی آن استفاده شده است [۱۷].
- روش یادگیری تقویتی عمیق DRL: روش یادگیری تقویتی مبتنی بر گرادیان با شبکه عصبی عمیق کاملاً متصل [۱۸].
- روش شبکه عصبی بازگشتی مانده با تکنیک عملگر-منتقد غیرسنکرون $A^{\pi}C-R^{\pi}N^{\pi}$: روشی که در آن از ساختار شبکه عصبی بازگشتی مانده و از تکنیک عملگر-منتقد غیرسنکرون برای آموزش استفاده شده

است

[۱۹]



شکل ۴. مقایسه روش ارائه شده با روش‌های مبتنی بر شبکه‌های عصبی

در این بخش به منظور مقایسه روش ارائه شده با روش‌های دیگر، نتایج روش ادغام ماشین‌های مجازی با امکان یادگیری اثرات ناهمبسته معیارهای انتخاب ماشین‌های مجازی برای مهاجرت (سناریو ۶ در قسمت قبل) نیز نشان داده شده است.

مقایسه روش پیشنهادی با سایر روش‌ها نشان می‌دهد که روش ارائه شده نتایج بهتری در مقایسه با سایر روش‌ها در تمامی معیارهای زمان پاسخ، نقض توافق سطح سرویس، مصرف انرژی و هزینه از خود نشان می‌دهد. علاوه بر این روش ارائه شده دارای قابلیت مقیاس‌پذیری است، برخلاف روش‌های مورد مقایسه که در ساختار شبکه عصبی خود تعداد ثابتی از میزبان‌ها در خروجی شبکه باید در نظر گرفته شود و مقیاس پذیر نیستند.

نتیجه‌گیری

در این پژوهش روش چندمعیاره جدیدی برای ادغام ماشین‌های مجازی در محیط ابری ارائه شد. کارهای قبلی نه تنها معیارهای محدودی را برای تصمیم‌گیری انتخاب ماشین‌های مجازی برای مهاجرت لحاظ می‌کنند، بلکه همبستگی معیارهای تصمیم‌گیری را نادیده می‌گیرند. این محدودیت‌ها منجر به سوگیری نسبت به انتخاب

گزینه‌هایی از ماشین‌های مجازی با ارزیابی خوب در دو یا چند معیار وابسته برای مهاجرت می‌شوند. به طور خاص، روش پیشنهادی از روش‌های تحلیل مولفه اصلی و تاپسیس برای استخراج اثرات ناهمبسته معیارهای تصمیم‌گیری و اجتناب از سوگیری در محاسبه راه‌حل‌های ایده‌آل (به دلیل همبستگی بین معیارها) استفاده می‌کند. نتایج تجربی نشان می‌دهد که پذیرش راه حل پیشنهادی منجر به کاهش ۴۱٫۵٪، ۳۰٫۱۳٪، ۱۲٫۹٪، ۱۳٫۴٪ و ۵۸٫۲٪ به ترتیب در مصرف انرژی، معیار نقض توافق سطح سرویس، زمان پاسخ، هزینه، تعداد مهاجرت‌های ماشین‌های مجازی در مقایسه با سایر روش‌ها می‌شود.

فهرست منابع

- [1] Zakarya, M., Gillam, L., Salah, K., Rana, O., Tirunagari, S., & Buyya, R. (2022). *CoLocateMe: Aggregation-based, energy, performance and cost aware VM placement and consolidation in heterogeneous IaaS clouds*. IEEE transactions on services computing, 16(2), 1023-1038. <https://doi.org/10.1109/TSC.2022.3181375>
- [2] Beloglazov, A., & Buyya, R. (2012). *Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers*. Concurrency and Computation: Practice and Experience, 24(13), 1397-1420. <https://doi.org/10.1002/cpe.1867>
- [3] Piraghaj, S. F., Dastjerdi, A. V., Calheiros, R. N., & Buyya, R. (2017). *ContainerCloudSim: An environment for modeling and simulation of containers in cloud data centers*. Software: Practice and Experience, 47(4), 505-521. <https://doi.org/10.1002/spe.2422>
- [4] Arianyan, E., Taheri, H., & Sharifian, S. (2015). *Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers*. Computers & Electrical Engineering, 47, 222-240. <https://doi.org/j.compeleceng.2015.05.006>
- [5] Rafiei, N., Amiri, F., & Mortazavi, H. (2023). *An Integrated Model to Evaluate the Design Concept Using the Analytic Hierarchy Process and TOPSIS Technique Based on Rough Numbers*. <https://doi.org/10.48301/KSSA.2023.375446.2367>
- [6] Yaghoubi, N., Dehghani, M., & Omidvar, M. (2017). *A model for the establishment of meta-synthesis technique based entrepreneurial university and TOPSIS*. Karafan Journal, 14(1), 51-65. <https://doi.org/20.1001.1.23829796.1396.14.41.3.4>
- [7] Horri, A., Mozafari, M. S., & Dastghaibyard, G. (2014). *Novel resource allocation algorithms to performance and energy efficiency in cloud computing*. The Journal of Supercomputing, 69(3), 1445-1461. <https://doi.org/https://doi.org/10.1007/s11227-014-1224-8>

- [8] Chen, T., Zhu, Y., Gao, X., Kong, L., Chen, G., & Wang, Y. (2018). *Improving resource utilization via virtual machine placement in data center networks*. *Mobile Networks and Applications*, 23(2), 227-238. <https://doi.org/https://doi.org/10.1007/s11036-017-0925-7>
- [9] Ma, Z., Shao, S., Guo, S., Wang, Z., Qi, F., & Xiong, A. (2020). *Container migration mechanism for load balancing in edge network under power Internet of Things*. *IEEE Access*, 8, 118405-118416. <https://doi.org/10.1109/ACCESS.2020.3004615>
- [10] Mahmoudian, M., Khorsand, R., & Ramezanpour, M. (2022). *A Prediction based Proactive Resource Provisioning Strategy for Multi-objective Workflow Scheduling in Cloud Computing*. *Karafan Journal*, 19(3), 529-551. <https://doi.org/10.48301/KSSA.2022.341879.2104>
- [11] Lebre, A., Pastor, J., Simonet, A., & Südholt, M. (2018). *Putting the next 500 vm placement algorithms to the acid test: The infrastructure provider viewpoint*. *IEEE Transactions on Parallel and Distributed Systems*, 30(1), 204-217. <https://doi.org/10.1109/TPDS.2018.2855158>
- [12] Wu, Q., Ishikawa, F., Zhu, Q., & Xia, Y. (2016). *Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters*. *IEEE transactions on services computing*, 12(4), 550-563. <https://doi.org/10.1109/TSC.2016.2616868>
- [13] Jiang, H.-P., & Chen, W.-M. (2018). *Self-adaptive resource allocation for energy-aware virtual machine placement in dynamic computing cloud*. *Journal of Network and Computer Applications*, 120, 119-129. <https://doi.org/10.1016/j.jnca.2018.07.011>
- [14] Pham, X.-Q., & Huh, E.-N. (2016). *Towards task scheduling in a cloud-fog computing system*. 2016 18th Asia-Pacific network operations and management symposium (APNOMS),
- [15] Gholipour, N., Arianyan, E., & Buyya, R. (2020). *A novel energy-aware resource management technique using joint VM and container consolidation approach for green computing in cloud data centers*. *Simulation Modelling Practice and Theory*, 104, 102127. <https://doi.org/10.1016/j.simpat.2020.102127>
- [16] Khan, A. A., Zakarya, M., Khan, R., Rahman, I. U., & Khan, M. (2020). *An energy, performance efficient resource consolidation scheme for heterogeneous cloud datacenters*. *Journal of Network and Computer Applications*, 150, 102497. <https://doi.org/10.1016/j.jnca.2019.102497>
- [17] Basu, D., Wang, X., Hong, Y., Chen, H., & Bressan, S. (2019). *Learn-as-you-go with megh: Efficient live migration of virtual machines*. *IEEE Transactions on Parallel and Distributed Systems*, 30(8), 1786-1801. <https://doi.org/10.1109/TPDS.2019.2893648>

- [18] Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource management with deep reinforcement learning. Proceedings of the 15th ACM workshop on hot topics in networks,
- [19] Tuli, S., Ilager, S., Ramamohanarao, K., & Buyya, R. (2020). *Dynamic scheduling for stochastic edge-cloud computing environments using a3c learning and residual recurrent neural networks*. IEEE transactions on mobile computing, 21(3), 940-954. <https://doi.org/10.1109/TMC.2020.3017079>
- [20] Li, W., Yue, H. H., Valle-Cervantes, S., & Qin, S. J. (2000). *Recursive PCA for adaptive process monitoring*. Journal of process control, 10(5), 471-486. [https://doi.org/10.1016/S0959-1524\(00\)00022-6](https://doi.org/10.1016/S0959-1524(00)00022-6)
- [21] Manimurugan, S. (2021). *IoT-Fog-Cloud model for anomaly detection using improved Naïve Bayes and principal component analysis*. Journal of Ambient Intelligence and Humanized Computing, 1-10. <https://doi.org/10.1007/s12652-020-02723-3>
- [22] Shen, S., Van Beek, V., & Iosup, A. (2015). Statistical characterization of business-critical workloads hosted in cloud datacenters. 2015 15th IEEE/ACM international symposium on cluster, cloud and grid computing,
- [23] Tuli, S., Mahmud, R., Tuli, S., & Buyya, R. (2019). *Fogbus: A blockchain-based lightweight framework for edge and fog computing*. Journal of Systems and Software, 154, 22-36. <https://doi.org/10.1016/j.jss.2019.04.050>