

International Journal of
Engineering Research and Science & Technology



ISSN:2319-5991

www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

INTELLIGENT CLOUD VM PLACEMENT FOR MINIMIZING IMAGE RETRIEVAL & COMMUNICATION OVERHEAD

Amar Gujeti, MSC in Computer Science and Cloud Computing, E-mail: amargujeti9@gmail.com

Abstract: This study represents a heuristic technique focused on optimizing the deployment of a virtual machine (VM) in cloud data centers focusing on minimizing image search and communication expenses. It presents three new algorithms: PM clustering, VM division and VM-PM mapping. The aim of the PM clustering is to reduce the overall communication operation of the longest communication distance between physical machines (PMS) and their aggregation to improve communication efficiency. The VM distribution uses a gross assignment strategy for clusters of analogous VM, increases image search performance and minimizes communication between VM. VM-PM mapping ensures optimal alignment of virtual machines with physical hosts according to the requirements for network and resources topology. Kmeans clustering is used to increase the location by taking into account the distance, search costs and communication. The AWS Cloud implementation shows the usability of real world contexts using EC2 instances and VPC architecture for testing and verification. This complete technique improves VM location, optimizes resource use and significantly reduces operating costs in cloud infrastructures.

“Index Terms - Virtual Machine Placement, AWS, Cloud Data Centers, PM Clustering, VM Partitioning, VM-PM Mapping, Resource Optimization, Image Retrieval, Communication Cost, Network Efficiency, Workload-Aware Allocation.”

1. INTRODUCTION

Infrastructure as a service (IAAS) has become a basic element of cloud computing, allowing users to install programs using virtual machines (VMS), which provide critical computing sources such as CPU, memory and storage. However, extensive deployment of VM often leads to considerable narrow places, especially inside storage systems inside the backend, where loading VM image files (VMIS) can lead to a serious network overload and degradation [14], [19]. The increasing need for

rapid VM image has caused research to increase VM tactics, facilitate faster securing and minimizing resource dispute [12], [18].

In addition to getting images of images, frequent communication across virtual machines can deepen the network load on the data center, leading to increased latency and reducing overall efficiency [5], [23]. Existing research has examined several optimization strategies for the location of VM, including the methods of consciousness and energy efficiency [1], [4], [24]. Numerous solutions use heuristic and meta-heuristic algorithms to increase the allocation of virtual machines while reducing energy consumption and communication expenditures [2], [5].

This article presents a heuristic approach to the location of VM, which is intended to improve the search for VMI and inter-VM communication in the architecture of the grid network. The solution uses the tripartite methodology: (i) PM clustering, which categorizes physical machines (PMS) according to the availability of sources and proximity of the network, (ii) VM division defined load and millional calculations. This method verified by AWS cloud architecture using EC2 and VPC services increases network efficiency, minimizes search delay and optimizes resource use in extensive cloud data centers.

2. RELATED WORK

The location of the virtual machine is a key element of cloud computing, which directly affects the use of resources, energy efficiency and network performance. A variety of methodologies have been examined to increase VM positioning techniques, emphasize aspects such as images of acquiring images, inter-VM communication and comprehensive scalability of the system.

Numerous research has examined the energy - efficient use of virtual machines to minimize the use of energy in cloud data centers. Parvizi and Resvani [4] have introduced a strategy for the location of the virtual machine using the meta-heuristic NSGA-III method to optimize the

allocation of workload and reduce energy consumption. Wei et al. [5] He proposed an increased approach to an Ant colony of optimization to increase energy efficiency while ensuring a suitable allocation of the virtual machine in data centers. These strategies seek to minimize energy consumption while maintaining the performance level.

Another area of study is to deploy a virtual machine that seeks to reduce overhead costs in the network as a result of operation between VM. Farzai et al. [3] Invented the optimization model with multiple lenses, which includes communication costs for VM allocation between data center servers. Meg et al. [23] They proposed a virtual machine location method that improves the scalability of the data center network by optimizing the location of the virtual computer according to the network topology. Zhao et al. [24] They have advanced this concept by concurrent optimization of the location of the virtual machine and a network design to increase the scalability of operation in dynamic cloud settings.

Extensive research has been carried out on effective VM image and storage optimization. Zhang et al. [19] They examined the collaborative into the VM image to speed up the initial duration of the VM and alleviate the storage systems. Alternative methodologies that proposed by Li et al. [17] and Yang et al. [20], use the similarity of content across VM images to increase storage and search efficiency to speed up the provision times. Similarly Xu et al. [14] He advocated a re-evaluation of the Methodologists to store VM images to improve search performance in extensive cloud implementations.

In addition, the investigation of migration and VM placement in the network dealt with the complexity of the living VM migration in a geographically distracted cloud setting. Al-Kiswany et al. [21] They introduced VMFlock, a co-migration method that increases the virtual machine transitions and minimizes migration costs. Darrous et al. [15] Introduced Nitro, a virtual machine management system, aimed at strengthening virtual machines inside scattered cloud infrastructures. Addya et al. [22] They introduced a virtual machine migration framework for multi-level applications and increased the selection of migration according to the availability of sources.

Regardless of this development, contemporary methodology often focuses on energy efficiency, communication optimization or images, although these elements do not move into the cohesive framework. This research seeks to fill this gap by creating the technique of placing a virtual machine based on heuristics, which at the same time optimizes the efficiency of images, communication and use of network resources inside the Data Center Data Center.

3. MATERIALS AND METHODS

The proposed system seeks to increase the allocation of virtual machines (VM) inside the cloud data centers using three new algorithms: PM clustering, VM division and VM-PM mapping. Together, these techniques improve resource allocation efficiency and at the same time reduce the costs associated with searching for images and transmission overheads.

PM clustering categorizes physical machines (PMS) into clusters according to sources, which facilitates effective administration of computational resources. Numerous research has examined the methodology of sources to increase the allocation of virtual computers and reduce the use of energy in data centers [4], [5]. VM location techniques correspond to the target of clustering PM by improving efficiency and reducing idle agents.

The VM division further categorizes VM according to the characteristics of the workload to ensure the optimal assignment of the resources. Previous studies have shown that the deployment of a virtual machine knowing workload increases the efficiency in data demanding [2], [3]. Methods of optimization with multiple lenses that are responsible for workload and communication have shown increased performance in cloud settings through strategic assignment by VM according to application behavior.

Mapping VM-PM intelligently allocates virtual machines to physical servers to increase the efficiency of the data center. This methodology integrates the virtual machine deployment techniques that are known, [23], [24] that reduce communication expenditures between VM and increase scalability. In addition, the cooperative methods of the VM image cache have shown to improve search efficiency [19], which emphasizes the need for intelligent mapping tactics that reduces the consumption of latency and bandwidth.

The system is evaluated on the cloud AWS using EC2, Amazon VPC and Elastic Block Store (EBS) to replicate and evaluate the location of VM inside the realistic environment. The research confirmed the efficiency of the heuristic models of the location of VM and dynamic algorithms of clumping workload in AWS settings to reduce costs and increase performance [1], [20].



“Fig.1 Proposed Architecture”

Illustration (Fig. 1) shows the strategy of deploying a VM using a method of improved KNN. It includes PMS performing operating systems and adapting VM. The procedure is guaranteed by the fair assignment of the VM grouping of PMS according to their geographical proximity. Comparable VMs are consolidated in the same PM to increase the use of bandwidth and communication efficiency. KNN allocates VM to PMS on the basis of proximity and minimal similarity. The image illustrates the Optimization of VM location for the speed and use of resources while reducing the network imbalance in the cloud computing context.

i) Dataset Collection:

The data file used to evaluate the proposed secure keyword search system includes a structured assembly of encrypted documents, metadata and indexed keyword files. The publicly accessible cloud storage data sets such as e-mail file ENRON and Wikipedia dump acts as the main sources of text data [11]. These data sets have diverse text content, which causes them to be suitable for assessing search efficiency and encryption.

The information is pre-processed to replicate the environment of cloud storage in the real world using encryption documents using AES and ECC encryption methods [21]. The keyword indexing is done using Grabled Bloom filters and Cuckoo hash [20]. The data file also includes synthetically

created search queries to evaluate latency and system accuracy.

This data file guarantees a comprehensive assessment of search efficiency, storage expenditure, and security while maintaining user privacy in keywords based on cloud-based keywords [1].

ii) New User Sign Up:

This module allows new users to create an account inside the system. Users are obliged to provide their personal information, including login, password and other necessary login information. After sending the form, the user's data is safely archived in the database and are registered in the system, ready to go to the following phases.

iii) User Login:

After successful registration, users can log in with their registered login information. This module verifies the user's identity by comparing the login information with the data contained in the database. After verification, the user is allowed to access the system functions and sent to the primary dashboard.

iv) Define VM & PM Parameters: This module allows users to determine the amount of VMS and PMS for simulation. The module allows users to configure the initial environment by marking the necessary amount of VM and PMS for testing and simulation goals. In addition, users can enter or create the necessary parameters, including search costs, communication costs and distance between VM and PMS.

v) Propose VM Placement: This module performs the proposed method of placing the virtual computer. It uses established properties of VMS and PMS, including distance and cost to determine the appropriate assignment of VM to PMS. The aim is to reduce the total costs by posting and communication costs and at the same time guarantee a fair distribution of VMS within the available PMS. The module generates mapping indicating the VM assignment to PMS according to the proposed technique.

vi) Extension Placement: This module uses the extension of the KMEANS clustering algorithm, which takes into account several criteria, such as the cost, search and communication for the placement of a virtual computer. Sorted PMS according to these parameters and then allocates VM the most suitable PM in each category. This improved placement technique is expected to



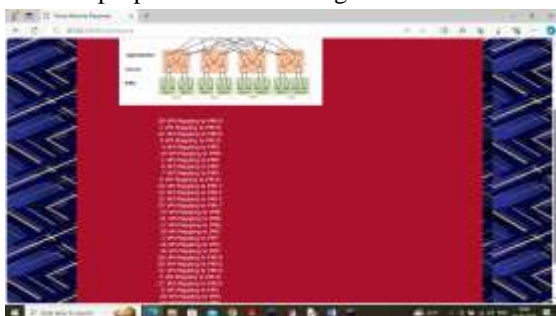
Click on the 'Define VM & PM Parameters' link above to specify the number of VMs and PMs used for the simulation.



On the aforementioned screen, input the values for VM and PM, then push the button to access the further page.



In the aforementioned screen, we are establishing distance, retrieval, and communication costs for each VM, since we lack actual PM and VM machines; hence, we are creating random values for distance, cost, and retrieval. Click on the 'Propose VM Placement' link to allocate the VM to the PM with the proposed method and get the result below.



The above screen displays the mapping of VM to PM using the proposed technique. Now, click on

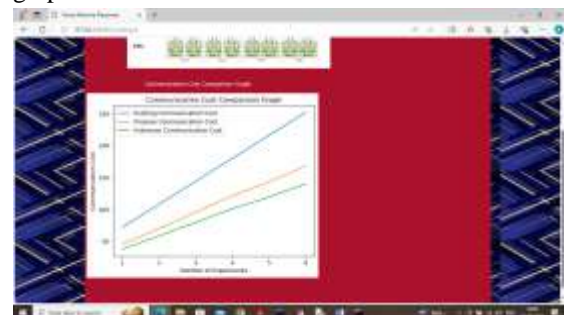
the 'Extension Placement' link to map VM to PM using the KMEANS extension algorithm, which will lead to the following page.



The following screen displays the mapping of each VM to its corresponding PM using the extension technique. Click on the 'Retrieval Graph' link to get the VM Image Retrieval graph below.



The graph's x-axis denotes the Number of Experiments, while the y-axis indicates 'Retrieval Cost.' The blue line illustrates the current cost, the orange line depicts the proposed cost, and the green line displays the extension idea. In all algorithms, the extension exhibited reduced retrieval time. Click on the 'Communication Cost' link to get the graph below.



The graph above illustrates the communication costs between VM and PM using current, proposed, and extended techniques. The x-axis of the graph denotes the 'Number of Experiments,' while the y-axis indicates the communication cost for each experiment. The blue line denotes the current method, the orange line signifies the proposed approach, and the green line illustrates the cost of extension placement.

Likewise, by adhering to the aforementioned screens, you may execute a simulation to allocate a virtual machine to a physical machine.

5. CONCLUSION

The proposed solution effectively reduces VM image search and communication expenditure in cloud data centers AWS, leading to an improvement in operating efficiency and a significant reduction in costs. Using heuristic algorithms-PM clustering, VM division and mapping VM-PM-System achieves an optimal and fair distribution of resources between the instances of Amazon EC2. This reduces the occurrence of overload and insufficient use and guarantees efficient use of AWS resources, such as the Elastic Block Store (EBS) and Virtual Private Cloud (VPC). Improvement of the system through KMEANS clustering, which takes into account factors such as network latency, storage duration and communication between VM, further specifies the accuracy of the selection of VM locations. This sophisticated optimization system reduces image and transmission expenses, while increasing performance with accelerated deployment and increased response time. For cloud service providers using AWS, this results in reduced operating costs and increased infrastructure scalability. End users obtain increased reliability, speedy service provision and more adaptive application environment. The system illustrates that the smart, cost-effective tactics of VM located inside a solid cloud platform, such as AWS, can significantly increase the scalability, efficiency and happiness of users in extensive cloud centers with high demand.

Future enhancements include implementation of VM dynamic migration for real-time workload changes, allowing continued optimization to optimize load fluctuations. The use of algorithms of predictive machine learning will improve resource forecasts and facilitate proactive allocation. Implementation of strict security protocols ensures the protection of personal data and compliance with the regulations in the settings of multiple tenants. In addition, the extension of the Hybrid cloud infrastructure system would provide optimal VM allocation across public clouds, private clouds and marginal devices, and therefore increased the overall performance, reliability and use of resources in different computing settings.

REFERENCES

- [1] S. Omer, S. Azizi, M. Shojafar, and R. Tafazolli, "A priority, power and traffic-aware virtual machine placement of IoT applications in cloud data centers," *J. Syst. Archit.*, vol. 115, May 2021, Art. no. 101996.
- [2] S. Sadegh, K. Zamanifar, P. Kasprzak, and R. Yahyapour, "A twophase virtual machine placement policy for data-intensive applications in cloud," *J. Netw. Comput. Appl.*, vol. 180, Apr. 2021, Art. no. 103025.
- [3] S. Farzai, M. H. Shirvani, and M. Rabbani, "Multi-objective communication-aware optimization for virtual machine placement in cloud datacenters," *Sustain. Comput., Inform. Syst.*, vol. 28, Dec. 2020, Art. no. 100374.
- [4] E. Parvizi and M. H. Rezvani, "Utilization-aware energy-efficient virtual machine placement in cloud networks using NSGA-III meta-heuristic approach," *Cluster Comput.*, vol. 23, no. 4, pp. 2945–2967, 2020.
- [5] W. Wei, H. Gu, W. Lu, T. Zhou, and X. Liu, "Energy efficient virtual machine placement with an improved ant colony optimization over data center networks," *IEEE Access*, vol. 7, pp. 60617–60625, 2019.
- [6] Y. Zhao, H. Chen, S. Zhao, and Y. Wang, "The storage of virtual machine disk image in cloud computing: A survey," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, 2017, pp. 263–267.
- [7] A. Greenberg et al., "VL2: A scalable and flexible data center network," in *Proc. ACM SIGCOMM Conf. Data Commun.*, 2009, pp. 51–62.
- [8] M. Mao and M. Humphrey, "A performance study on the VM startup time in the cloud," in *Proc. IEEE Fifth Int. Conf. Cloud Comput.*, 2012, pp. 423–430.
- [9] K. Jayaram, C. Peng, Z. Zhang, M. Kim, H. Chen, and H. Lei, "An empirical analysis of similarity in virtual machine images," in *Proc. Middleware Ind. Track Workshop*, 2011, pp. 1–6.
- [10] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 63–74, 2008.
- [11] C. Peng, M. Kim, Z. Zhang, and H. Lei, "VDN: Virtual machine image distribution network for cloud data centers," in *Proc. IEEE INFOCOM*, 2012, pp. 181–189.

- [12] J. Reich et al., “VMTorrent: Scalable P2P virtual machine streaming,” in Proc. Int. Conf. Emerg. Netw. Exp. Technol., vol. 12, 2012, pp. 289–300.
- [13] Z. Zhang et al., “VMThunder: Fast provisioning of large-scale virtual machine clusters,” IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 12, pp. 3328–3338, Dec. 2014.
- [14] X. Xu, H. Jin, S. Wu, and Y. Wang, “Rethink the storage of virtual machine images in clouds,” Future Gener. Comput. Syst., vol. 50, pp. 75–86, Sep. 2015.
- [15] J. Darrous, S. Ibrahim, A. C. Zhou, and C. Perez, “Nitro: Network-aware virtual machine image management in geo-distributed clouds,” in Proc. IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID), 2018, pp. 553–562.
- [16] M. Björkqvist, L. Y. Chen, M. Vukolic, and X. Zhang, “Minimizing retrieval latency for content cloud,” in Proc. IEEE INFOCOM, 2011, pp. 1080–1088.
- [17] H. Li, W. Li, Q. Feng, S. Zhang, H. Wang, and J. Wang, “Leveraging content similarity among VMI files to allocate virtual machines in cloud,” Future Gener. Comput. Syst., vol. 79, pp. 528–542, Feb. 2018.
- [18] H. Li, S. Wang, and C. Ruan, “A fast approach of provisioning virtual machines by using image content similarity in cloud,” IEEE Access, vol. 7, pp. 45099–45109, 2019.
- [19] Y. Zhang, K. Niu, W. Wu, K. Li, and Y. Zhou, “Speeding up VM startup by cooperative VM image caching,” IEEE Trans. Cloud Comput., vol. 9, no. 1, pp. 360–371, Jan.–Mar. 2021.
- [20] Y. Yang, B. Mao, H. Jiang, Y. Yang, H. Luo, and S. Wu, “SnapMig: Accelerating VM live storage migration by leveraging the existing VM snapshots in the cloud,” IEEE Trans. Parallel Distrib. Syst., vol. 29, no. 6, pp. 1416–1427, Jun. 2018.
- [21] S. Al-Kiswany, D. Subhraveti, P. Sarkar, and M. Ripeanu, “VMFlock: Virtual machine co-migration for the cloud,” in Proc. Int. Symp. High Perform. Distrib. Comput., 2011, pp. 159–170.
- [22] S. K. Addya, A. Satpathy, B. C. Ghosh, S. Chakraborty, and S. K. Ghosh, “Power and time aware vm migration for multi-tier applications over geo-distributed clouds,” in Proc. IEEE Int. Conf. Cloud Comput. (CLOUD), 2019, pp. 339–343.
- [23] X. Meng, V. Pappas, and L. Zhang, “Improving the scalability of data center networks with traffic-aware virtual machine placement,” in Proc. IEEE INFOCOM, 2010, pp. 1–9.
- [24] Y. Zhao, Y. Huang, K. Chen, M. Yu, S. Wang, and D. Li, “Joint VM placement and topology optimization for traffic scalability in dynamic datacenter networks,” Comput. Netw., vol. 80, pp. 109–123, Apr. 2015.
- [25] O. Biran et al., “A stable network-aware VM placement for cloud systems,” in Proc. IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID), 2012, pp. 498–506.