A Comparative Analysis of Virtual Machine Placement Techniques in the Cloud Environment

Bhavesh Gohil
S.V.National Institute of
Technology, Surat
Gujarat, India

Sanjana Shah
S.V.National Institute of
Technology, Surat
Gujarat, India

Yash Golechha
S.V.National Institute of
Technology, Surat
Gujarat, India

Dhiren Patel
S.V.National Institute of
Technology, Surat
Gujarat, India

ABSTRACT
Cloud computing is a novel paradigm that aims to provision on-demand computing capacities as services. Virtualization is an important technology integrated in Cloud Computing. Mapping the virtual machines to the appropriate physical machines is called VM placement. The effectiveness and elasticity of virtual machine placement has become the main concern in cloud computing environments. Effective placement of virtual machines is important for optimization of computational resources and reduction of the probability of virtual machine reallocation. This paper provides a survey and brief analysis of some of the main VM Placement mechanism utilized in cloud computing.

General Terms
Virtual Machine, Virtual Machine Placement, Cloud Computing

Keywords
Virtual Machine Placement, Constraint Programming, Stochastic Integer Programming, Bin Packing, Genetic Algorithm, Cloud computing, Performance evaluation

1. INTRODUCTION
The necessary growth in the need for pervasive computing utilization models has resulted in the realization of a new computing model called cloud computing in which consumers from different locations and devices can access shared computing resources in an on-demand, pay-as-you go model. [2] The cloud computing model entails an adaptive resource allocation management model to handle the workload of several applications with different resource allocation requirements. Additionally, to accommodate different requests of numerous applications simultaneously for improved quality of service and data centers, using Virtual Machines (VMs) over Physical Machines (PMs) for resource consolidation and environment isolation is one probable solution. In cloud computing, VM placement is a critical task that is conducted as part of the VM migration process. It aims to find the best and most suitable PM to host the VMs. Many VM placement algorithms have been designed and proposed in the cloud computing environment to improve various factors affecting the data centers, VMs and their executions. Some of these algorithms are only suited for specific environments and circumstances so it is necessary to analyze the advantages and disadvantages of each, to thoroughly understand which type of VM Placement scheme is appropriate for which situation. The remainder of this paper is organized as follows. Section 2 describes literature survey and theoretical background of virtual machine placement, and classification of its algorithms via major approaches currently popular in the industry today. Sections 3 to 6 present the 4 main virtual machine placement techniques and approaches in the cloud environment. Sections 7 and 8 compare the mentioned approaches based on performance metrics and evaluate their strength in different situations and section 9 present the concluding remarks.

2. VIRTUALIZATION ARCHITECTURE
2.1 Virtualization Architecture
The IaaS service model supports resource multiplexing through virtualization technology which offers a means to decouple the application activities from the physical resources required. This decoupling of resources is facilitated by the concept of a 'virtual machine' which compresses an application with a specific set of functionalities. In a virtualized environment, numerous applications run on a VM and then one or more VMs are mapped onto each PM of the cloud data center. [5] The VM placement is a critical process which is conducted to determine the most appropriate PM or server to host the VM. Selecting a suitable host is very important to improve power efficiency, resource utilization and QoS support in a cloud computing environment.

2.2 Classification of VM Placement(VMP) Algorithms
Depending on the goal of placement, a VM placement algorithm can be categorized into two types:

A) Power-based VMP: algorithms obtaining a VM-PM mapping which allows a system to be energy-efficient and procure utmost resource utilization.

B) QoS-based VMP: algorithms obtaining a VM-PM mapping to ensure and satisfy of quality of service requirements

Depending on the necessity of migration, (transferring a VM from one PM to another PM) a VM Placement algorithm can be classified into 2 types:

A) Static VMP- VM placement does not consider either the states of the virtual machines and physical

machines, or the arrival rate of the user requests.

B) Dynamic VMP- achieve optimum solutions from the already present mapping of VMs at minimal cost.

Furthermore, the dynamic VM placement algorithms can be categorized as reactive and proactive VM placement.

a) Proactive VMP: changes the VM's Physical Machine of an initial placement before the system reaches a certain condition or satisfies a specific criteria.

b) Reactive VMP: changes to an initial placement after the system reaches a certain undesired state. The change to this original placement could be because of the performance, maintenance, power or load problems or some SLA violations.

3. CONSTRAINT PROGRAMMING BASED VMP

Constraint programming is a programming technique that is applied for VM placement in cloud computing environments. It is beneficial for combinatorial search problems where the solutions must satisfy the constraints on relations between variables. [21]A set of constraints are set and they can be easily extended further to involve and satisfy more requirements. Cloud service providers aim to automate management of virtual machines by taking the quality of service requirements of applications into consideration. This can be represented as a constraint programming problem. The primary goal of the constraint programming is to maximize a global utility function. This global utility function is chosen keeping in mind SLA fulfilment and operating costs. It maps the current state of each application (workload, resource capacity, SLA) to a scalar number.[5] This scalar value tries to quantify the applications' satisfaction with respect to the goals that are set by the automatic manager.

The virtual machine placement using a constraint based approach can be thought of as a two-stage process.

Stage 1: Local Decisions, associated with each application environment.

Stage 2: Global Decisions, that takes as input the local decision from all applications, and then tries to maximize the global utility function.

Any application hosted in the cloud is called an application environment (AE). [11] Assume that one VM is associated with only one AE. A Local decision module is associated with each AE which computes a utility function. This utility function gives us a measure of the application satisfaction with a specific resource allocation. This output is given to a global decision module which maps the VMs to PMs considering the CPU load of virtual and physical machines. At the Local Decision Module, decisions are associated with a specific application environment. It takes into account a fixed service-level utility function mapping the service level to a utility value and a dynamic resource-level utility function mapping a resource capacity to a utility value. The result of resource-level utility function mapping is communicated to Global Decision Module at every iteration.

There are 2 constraints in Stage 1: First, the number of VMs of a specific class has to meet the capacity requirement of the PM. Second, each application associates an upper bound on

the number of VMs of each class and the total number of VMs that the PM is willing to accept.[3] Again there are constraints at Stage 2: First, the VMs allocated to all the applications are bound by CPU and RAM capacity constraints on the physical servers or hosts. The ultimate goal is to reduce the number of physical machines or active number of hosts.

4. STOCHASTIC INTEGER PROGRAMMING BASED VMP

Stochastic integer programming (SIP) is applied for modelling optimization problems that involve a high amount of uncertainty. [1] Stochastic means "having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely." SIP is useful in cases where actual demands are not known but the distribution of demands is known or can be estimated. In SIP, there are 3 phases in which to allocate resources:

A) Reservation: Cloud broker provisions resources and estimates the need without knowing the demand of users

B) Utilization: The reserved resources are actually used by the cloud service users.

C) On-demand: In case the demand exceeds the reserved resources, the additional resources can be requested in an on-demand pay plan.

SIP is used to rent resources from providers and measure the trade-off between the advance reservation of resources, and the optimal allocation of on-demand resources. [4]This scheme consists of two stages:

Stage 1: Defines the number of the VMs provisioned in the reservation phase.

Stage 2: Defines the number of the VMs allocated in both utilization and on-demand phases. [14] (In other words, the second stage represents the actual number of the VMs required by the user and actual prices defined by providers.)

The problem statement has the following:

Stage 1: The set of VM classes, application types, and cloud providers, resources in all 3 phases, maximum capacity, and cost along with demands are stated.

Stage 2: The SIP Formulation – the correct mapping is found and selected for the placement.

Stochastic Integer Programming is helpful in estimating the variation in demand and costs, thereby, frequent re-computations are not needed. [2] However, this means if there is error in the estimation, users might end up paying more.

5. BIN PACKING BASED VMP

The bin packing-based VM can be used to find the actual mapping of virtual machines to available physical machines. It is possible to minimize the cost of running data centers by tightly packing the VMs required to be running at a time onto the least number of PMs possible. [6] However, care should be taken so as to avoid overloading which results in SLA violations.

The individual physical machines can be considered as bins having different dimensions. These dimensions correspond to resource capacities of the physical machines, in the situation. Similarly, the virtual machines can be considered as objects to be packed into these bins or PMs. Again, for each VM, the amount of resources required (dimensional requirements of

objects) is specified. [9] The bin packing algorithm is used to find a mapping between these objects (VMs) and bins (PMs) such that the total number of bins required is minimized. The Bin Packing problem may be a good tool to model the VM deployment problem. However a PM has multiple computing resources and the conventional Bin Packing problem cannot model multiple resources. Thus, the Multi-Dimensional Bin Packing Problem (MBP) should be used for those cases. The MBP can be formulated using a mixed-integer linear programming (MILP) model. [13] The main inputs to the algorithms are the resource requirements of each virtual machine to be placed. 2 matrices for requirements and allocation for the virtual machines and their dimensions are formed. Given the resource requirements of virtual machines, an algorithm is expected to come up with a placement plan which gives a mapping of which virtual machine is to be placed on which physical machine. This placement plan given by an algorithm can be captured in the form of an allocation matrix. To solve this problem, one essentially has to calculate this allocation matrix. [12] For this, one implements an LP (Linear Programming) formulation of the problem and also define some heuristics.

In order to solve this problem using Integer Linear Programming, the constraints that need to be satisfied during placement of virtual machines have to be stated. They are:

A) Capacity Constraints: For each dimension of a given PM j, the sum of the resource requirements of all VMs placed on it should be less than or equal to the total available capacity.

B) Placement Guarantee Constraints: All virtual machines must be placed.

5.1 First Fit

Based on the forecast of demand of each VM, the VMs are sorted in descending order. Each VM is then taken off the list and an attempt is made to place it on the first PM where it fits, that is, after its placement, the sum of resource demands of all VMs on that PM does not exceed the total capacity of the PM. [16]The first-fit algorithm minimizes the number of active PMs. The aspect of interest for us is the reduction in number of used PMs versus the SLA violations. In comparison to the number of PMs used for static allocation, the number of PMs used in dynamic allocation is reduced to half. For a given SLA violation rate, the average number of PMs used by the above algorithm is less than the static algorithms. When the first-fit approximation is used, for finding a VM to PM map, the migration costs should also be taken into account. Since, the algorithm uses first approximation, it might fail to deliver the optimal solution as all possible solutions in the solution space would not be checked.[15] The placement algorithm is dependent on the efficiency of the forecasting technique. If the error in the forecast is huge, the number of SLA violations can go beyond the accepted value. This method does not take constraints other than capacity of resources into account. Thus it may end up putting two interfering VMs on one PM.

5.2 First Fit Decreasing

Both physical machines and virtual machines are sorted from high capacity to least capacity machines. Once both physical and virtual machines are arranged, first fit process takes place. Time complexity of this algorithm is O (n+m). When the group delivery mode is employed, the First Fit Decreasing Algorithm, a variant of the First Fit Algorithm, can further reduce the average number of active hosts. The primary difference is that the set of virtual machines in the queue is ordered according to some criteria before placement. In a simple model, where just one dimension is considered (typically, CPU usage), [7] virtual machines are ordered according to their demand for the corresponding resource. In a more sophisticated model, several dimensions can be considered to establish a ranking amongst the virtual machines. [8] One way of defining the rank of each virtual machine is to employ the volume method whereby the volume of a virtual machine is calculated by multiplying its demands in all dimensions.

5.3 Best Fit Decreasing

The Best-fit Decreasing algorithm first sorts the VMs in the decreasing order of its VM utilization and for each host in the host list, if the host has enough resource for the VM, then selects the host as a destination, otherwise it does nothing. The best-fit algorithm is resource-aware and energy-aware, but not traffic-aware, cost-aware, or application-aware. Along with saving energy, a data center should try to fulfil all SLA requirements, including virtual machine placement ratio, performance and elasticity. The Best Fit Algorithm keeps a set of hosts active at all times and places each VM at the best position to achieve load balance amongst the active hosts. [11]This should increase the probability of success in VMPs. It should also increase the probability given that the host is fulfilling the needs due to a virtual machine expansion. Thus, virtual machine migration is reduced, preventing systems performance degradation.

5.4 Worst Fit

In the Worst-fit algorithm, the cloud broker places the VMs to less active servers and restricts the mapping of the VMs, such that power consumption and computation cost are minimized. [17]In addition, the first PMs and VMs are sorted in the decreasing order of utilization. Then, the first PM which has the required resources from the list of the sorted PMs is selected and then deployed. This procedure is repeated till all the VMs are mapped to the PMs.

5.5 Next Fit

In this algorithm, there will be a variable name called next. At Initial, next is null. First virtual machine search starts from the first node.[14] When physical machine satisfies the resource requirement, virtual machine starts on that physical machine and next is replaced by current physical machine. If there is no PM with sufficient available resources, a new PM will start up and create the requested VM on the newly started PM. For the next virtual machine search starts from next to the stored next physical machine, search continues up to desired physical machine.

5.6 Random Fit

Randomly a physical machine is chosen for placing virtual machine. [18] If physical machine satisfies the resource requirement, virtual machine starts on that physical machine. If it not satisfied randomly another physical machine is chosen. This process will continue for every virtual machine.

6. GENETIC ALGORITHM

The Genetic Algorithm (GA) is a population-based metaheuristic, which uses evolution to find better mappings between VMs and PMs. Evolutionary algorithms use techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover to generate solutions to optimization problems.[19] Usually the evolution starts from random initial population, and then in generation, the specimens are modified with the use of genetic operators and

their fitness is evaluated in order to select best solutions for the next generation.

The GA is a heuristic based search technique. It can be called as bin packing extended with additional constraints. It is particularly useful in problems where objective functions dynamically change.

The GA mainly involves:

A) Chromosome modelling - This is an encoding schema used to encode details of the problem that is to be passed from one generation to the other.

B) Population Initialization - The feasibility of the solution is dependent on the feasibility of the initial solution.

A) Crossover - It is a genetic operator used to vary the programming of a chromosome from one generation to next.

B) Mutation - Mutation is a genetic operator that alters one or more gene values in a chromosome from its initial state. This helps to prevent the population from stagnating at any local optima. [21]

C) Generation Alternation - This is where one will select one of the solutions from the next generation of solution.

GA focuses on multiple objectives: resource wastage minimization, power consumption minimization and cost of thermal dissipation minimization. [3] Simulation proves that the various GA-based algorithms solve the conflicting objectives while the bin packing algorithms cannot.

7. PERFORMANCE METRICS

Automating the process of virtual machine placement has become important with the increase in size of the data centers. Some of the factors that are typically used to quantify the cost functions are:

1. Resource usage (CPU, Memory, I/O, RAM, etc.)

2. Network Usage

3. Power and Energy Consumption

4. Cost

The choice of which performance factors and how many performance factors are to be taken into consideration varies from cloud provider to cloud provider.

7.1 Resource Usage

Virtual resource management is one of the important issues in cloud computing. Normally, each VM may access various resources to execute its applications.[21]Resource-aware VM placement schemes consider hardware resource requirements of the VMs in the placement decisions. An efficient resource-aware placement scheme tries to optimally place VMs on the PMs such that the overall resource utilization is maximized.

7.2 Network Usage

The intra-data and inter-data center network traffic has a significant effect on the SLAs, revenue of the cloud providers and the performance of the cloud computing services. With the increasing trend towards communication applications, network-aware VM placement schemes have become very important. These schemes address the traffic related issues in the VM placement for the cloud data centers and try to reduce network traffic. [20] Also, some of these schemes distribute network traffic evenly and try to avoid congestion.

The following factors play an important role in network-aware VM placement schemes:

1. Minimizing the data transfer time consumption: Optimizing the data transfer between the virtual machine and data.

2. Traffic between the VMs: Decreasing the overall network traffic and minimizing the average path length between the VMs, improve communication performance by reducing, minimizing the congestion of network between the VMs.

3. Traffic between remote cloud sites: Reducing traffic in the federated clouds.

4. Traffic between the PMs.

5. Distance between the VMs.

7.3 Power and Energy Consumption

Due to the high cost of power consumption as well as the concerns for global warming and CO2 emissions, recently, the green data centers or "green cloud" have become an important issue for cloud service providers. Power-aware VM placement schemes are very important in green cloud computing which tries to make cloud data centers more efficient and reduce the power consumption of the data centers. [2] This results in lower maintenance costs for the cloud provider. The following factors are taken into consideration when determining the power and energy consumption of a VM Placement Schema.

1. CPU Utilization: The state of being idle or busy according to the utilization of processors. (Four defined status explain as Idle, Average, Active, Over-utilized)

2. Server States: The amount of power that various server components (CPU, memory, storage) consumes.

3. PM States: Different states of PM, consume different amounts of power- (4 states: Running, Ready, Sleep, Off)

4. Network Elements: Reducing network routing elements' costs.

5. Data center power usage: The cost is divided into four categories: server base energy, server dynamic energy, cooling energy and peak power.

6. Distance between the VMs: decreasing distance between the VMs to save energy.

7.4 Cost

Cost-aware VM placement schemes try to mitigate the data centers maintenance costs for the cloud providers. The important issue in these schemes is to achieve the cost saving while considering the QoS of cloud services and honoring the SLAs. [3]The following factors play an important role in cost-aware VM placement schemes

A) VMs cost: The cost of powering the VMs when they are instantiated in the data centers.

B) Physical Machine Cost: The cost of using the PM in the specific period of time.

C) Data center cost.

D) Distance between the VMs and the clients: By reducing the network distance between the VMs and the clients the perceived user performance is improved.

E) Cooling cost: The cost of the cooling system that is used for the data centers.

8. COMPARISON OF VMP TECHNIQUES

Constraint Programming is useful in cases where input data is with us, that is, before computing the cost functions, one knows the demands of the Virtual Machines. The techniques using constraint programming are easily extendable so as to take additional constraints into account. It is easy to specify constraints so as to handle the trade-off between multiple objectives which might conflict with one another. The time taken to generate an optimal solution is high as the number of constraints increases.

With increase in ϵ (tradeoff factor) as the resource demands increases, the global utility value decreases.[14] The operating cost of the application is boosted by this factor. Thus an application having a higher weight might end up not getting the resources, because of the higher trade off factor. The application weight denotes the resource allocation priority. Therefore, an application with a higher weight will get priority over an application having a lower weight. These weights should be critically chosen. Thus, in a time interval where demands for two applications are high, the resources will be provisioned to applications having a higher weight. The other application might have to face SLA violations.[20]

Stochastic Integer Programming is useful where the future demands and prices of resources are not known, but their probability distributions are either known or can be computed. This is the best technique to be used in the case when there are two or more uncertain parameters on which the cost depends. [1] Stochastic Integer Programming is helpful in estimating the variation in demand and costs, thereby, frequent re-computations are not needed. However, this means if there is error in the estimation, users might end up paying more.

The Bin-packing approach can be considered as a subset of Constraint Programming approach. [4] It is useful for dynamic VM Placement, especially where the demand is highly variable. It is a heuristic based approach. Thus it may not give us an optimal solution. [12] However, it will always generate a good solution in considerable amount of time. The Bin-packing approach is really useful when all physical machines have the same amount of memory and processing capabilities. However, it is possible to model bin packing algorithms with constraints. There is always a tradeoff between packing maximum number of virtual machines on a single physical machine versus distributing the load across all the physical machines. Multi-dimensional bin-packing algorithm can also be applied to VM placement problem. In this the dimensions correspond to amount of memory and number of processing units.

Genetic Algorithm is a way to solve the bin packing problem with certain constraints. GA requires more computing time and higher computing resources as compared to bin packing. It is particularly useful for static placements, that is, in scenarios where the demands do not vary over a considerable period of time.[7] GA is also useful for specifying VM-VM and VM-PM interference constraints. GA works well in most of the cases, irrespective of the number of constraints being high or low. It is important to choose a feasible initial solution, because the genetic operators are applied to this solution.[11] In order to get a solution in fixed time, one can put a restriction on the number of generations. Thus, one may end up not getting the optimal solution, but the solution will be better than the one originally started off with.

Each of the virtual machine placement algorithm works well under certain specific conditions. Thus, it is important to choose a technique that suits the needs of the cloud user and cloud provider. Also, the parameters to these algorithms should be properly specified. [18] The performance metrics are measured at both system level and application level. The system level metrics are measured in terms of CPU load and the application level metrics are measured in terms of response time of applications. Table 1 presents the key differences between various virtual machine placement techniques.

9. CONCLUSION AND FUTURE SCOPE

Virtual Machine Placement in data centers of the cloud environment has been an active area of research in the past few years. This report shed light on various placement algorithms and their performance evaluation. The objective of these techniques can either be minimization of power consumption or providing QoS.[15] Ranking these algorithms or stating the best one is not a proper suggestion because every placement technique has some specific target, migration technique, prominent resources and influential parameters. Although these techniques may seem accurate and applicable externally, there exists some or the trade-off when surveyed thoroughly. Owing to the workload variability and continuously changing demands of applications, there is a need to constantly optimize these VM placement algorithms. A wrong VM deployment algorithm puts a VM on an unsuitable PM or deploys many VMs to a limited number of PMs. This may result in low resource utilization and load imbalance. It is essential that the right algorithm is used at the right place to ensure optimization, efficiency, and affordable costs.

As a future work introduction of fuzzy algorithms could take the advantages from different selection criteria and form a rule base for VM selection. Also, there arises the possibility of making more ecofriendly IT infrastructures with reasonable amount of on-demand operating cost to improve the quality of IaaS of cloud computing. Additionally, with the ever increasing security attacks on the VMs and the cloud services, providing security in the cloud environment has become a more critical issue. For example, many attacks such as VM Escape attacks, VM Sprawling attacks, Cloud-Internal Denial of Service attacks (CIDoS), VM Neighbor attacks etc. are conducted in the cloud computing that target the VMs and a majority of them force the system to migrate the VMs. By misusing the VM migration feature of the cloud, these attacks incur high overheads to the cloud system and degrade its performance and availability. As a result, security is one of the crucial factors which should be considered in the future VM placement researches and studies, and its negligence makes the whole cloud system vulnerable to various security attacks and hinder the cloud computing critical role in the future IT systems.

Table 1: Comparison of VMP Techniques

No.	Metric	Constraint Programming	Stochastic Integer Programming	Bin-packing	Genetic Algorithm
1	Input Data	Known	Not Known	Not Known	Known
2	Constraint Extension	Yes	No	No	Yes
3	Response Time	Not Optimum	Optimum	Not Optimum	Optimum
4	Static/Dynamic	Static	Dynamic	Dynamic	Dynamic
5	Technique Used	Deterministic	Deterministic	Heuristics	Meta- Heuristics
6	Actual Mapping	Yes	No	Yes	Yes
7	PM's allotted	Not Optimum	Not Optimum	Optimum	Optimum
8	Application	When input data is known to us	When probability distributions of resources can be calculated	To find the actual mapping of virtual machines to available physical machines	To solve server consolidation problem with conflicts.
9	Merits	Changes in the global utility function can be made at run time	Frequent re computations are not needed	The number of PMs used in dynamic allocation is reduced to half	Works well irrespective of number of constraints.
10	Demerit	Response time increases as number of constraints increases	It does not do the actual mapping of VMs to PMs	The solution is not best optimal solution as all solution space is not checked	Restriction on the number of generations.

10. REFERENCES

[1] Zhou, Z., Hu, Z., & Li, K. (2016). Virtual Machine Placement Algorithm for Both Energy-Awareness and SLA Violation Reduction in Cloud Data Centers. Scientific Programming, 2016, 1-11. doi:10.1155/2016/5612039

[2] Calcavecchia, N. M., Biran, O., Hadad, E., & Moatti, Y. (2012). VM Placement Strategies for Cloud Scenarios. 2012 IEEE Fifth International Conference on Cloud Computing. doi:10.1109/cloud.2012.113

[3] Thiruvenkadam, T., & Kamalakkannan, P. (2015). Energy Efficient Multi-Dimensional Host Load Aware Algorithm for Virtual Machine Placement and Optimization in Cloud Environment. Indian Journal of Science and Technology, 8(17). doi:10.17485/ijst/2015/v8i17/59140

[4] Zhou, Z., Hu, Z., Song, T., & Yu, J. (2015). A novel virtual machine deployment algorithm with energy efficiency in cloud computing. Journal of Central South University, 22(3), 974-983. doi:10.1007/s11771-015-2608-5

[5] Ma, F., Liu, F., & Liu, Z. (2012). Distributed load balancing allocation of virtual machine in cloud data center. 2012 IEEE International Conference on Computer Science and Automation Engineering. doi:10.1109/icsess.2012.6269396

[6] Sheikhalishahi, M., Wallace, R. M., Grandinetti, L., Vazquez-Poletti, J. L., & Guerriero, F. (2016). A multi-dimensional job scheduling. Future Generation Computer Systems, 54, 123-131. doi:10.1016/j.future.2015.03.014

[7] Amarante, S. R., Roberto, F. M., Cardoso, A. R., & Celestino, J. (2013). Using the Multiple Knapsack Problem to Model the Problem of Virtual Machine Allocation in Cloud Computing. 2013 IEEE 16th International Conference on Computational Science and Engineering. doi:10.1109/cse.2013.77

[8] Bin-Packing. (n.d.). SpringerReference. doi:10.1007/springerreference_5277

[9] Malviya, P., Agrawal, S., & Singh, S. (2014). An Effective Approach for Allocating VMs to Reduce the Power Consumption of Virtualized Cloud Environment. 2014 Fourth International Conference on Communication Systems and Network Technologies. doi:10.1109/csnt.2014.121

[10] Ni, J., Huang, Y., Luan, Z., Zhang, J., & Qian, D. (2011). Virtual machine mapping policy based on load balancing in private cloud environment. 2011 International

Conference on Cloud and Service Computing. doi:10.1109/csc.2011.6138536

[11] Hamdi, K., & Kefi, M. (2016). Network-aware virtual machine placement in cloud data centers: An overview. 2016 International Conference on Industrial Informatics and Computer Systems (CIICS). doi:10.1109/iccsii.2016.7462398

[12] Chowdhury, M. R., Mahmud, M. R., & Rahman, R. M. (2015). Study and performance analysis of various VM placement strategies. 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). doi:10.1109/snpd.2015.7176234

[13] Masdari, M., Nabavi, S. S., & Ahmadi, V. (2016). An overview of virtual machine placement schemes in cloud computing. Journal of Network and Computer Applications, 66, 106-127. doi:10.1016/j.jnca.2016.01.011

[14] Li, X., Qian, Z., Lu, S., & Wu, J. (2013). Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data center. Mathematical and Computer Modelling, 58(5-6), 1222-1235. doi:10.1016/j.mcm.2013.02.003

[15] Li, K., Zheng, H., & Wu, J. (2013). Migration-based virtual machine placement in cloud systems. 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet). doi:10.1109/cloudnet.2013.6710561

[16] Usmani, Z., & Singh, S. (2016). A Survey of Virtual Machine Placement Techniques in a Cloud Data Center. Procedia Computer Science, 78, 491-498. doi:10.1016/j.procs.2016.02.093

[17] Pires, F. L., Melgarejo, E., & Baran, B. (2013). Virtual machine placement. A multi-objective approach. 2013 XXXIX Latin American Computing Conference (CLEI). doi:10.1109/clei.2013.6670671

[18] Choudhary, A., Rana, S., & Matahai, K. (2016). A Critical Analysis of Energy Efficient Virtual Machine Placement Techniques and its Optimization in a Cloud Computing Environment. Procedia Computer Science, 78, 132-138. doi:10.1016/j.procs.2016.02.022

[19] Zheng, X., & Cai, Y. (2014). Dynamic Virtual Machine Placement for Cloud Computing Environments. 2014 43rd International Conference on Parallel Processing Workshops. doi:10.1109/icppw.2014.28

[20] Mills, K., Filliben, J., & Dabrowski, C. (2011). Comparing VM-Placement Algorithms for On-Demand Clouds. 2011 IEEE Third International Conference on Cloud Computing Technology and Science. doi:10.1109/cloudcom.2011.22.

[21] Zeng, H., Zeng, H., He, T., & Zhang, N. (2013). Virtual machine placement and optimization for data center. Journal of Computer Applications, 33(10), 2772-2777. doi:10.3724/sp.j.1087.2013.02772