

# Research on Virtual Machine Cluster Deployment Algorithm in Cloud Computing Platform

Zheng Yao, Wen-Sheng Tang<sup>(✉)</sup>, Sheng-Chun Wang, and Hui Peng

The Department of Computer Teaching, Hunan Normal University, Changsha, 410081, China  
154976552@qq.com

**Abstract.** To address the virtual machine cluster deployment issues in cloud computing environment, a novel MCSA (Min-cut segmentation algorithm) of virtual machine cluster is proposed with resource and communication bandwidth constraints. In this paper, the basic idea is based on the fully consideration on the CPU, memory, hard-disk and other resource constraints between virtual machine cluster and physical host, as well as the communication bandwidth constraints between the virtual machine. We quantified the virtual machine cluster constructed an undirected graph. In the undirected graph, the nodes represent the virtual machine, so the weight of a node represents the value of resources, and the edges represent the communication bandwidth, so the weight of the edge represents the value of communication bandwidth. Base on the above transformations, the resources and bandwidth constrained optimization problem is transformed into the graph segmentation problem. Next we segment the undirected graph by minimum cut algorithm, and computing the matching degree of physical machines. Last we obtained the approximate solution. To validate the effectiveness of the new algorithm, we carried out extensive experiments based on the CloudSim platform.

**Keywords:** Cloud computing · Virtual machine cluster deployment · Graph partitioning · Communication bandwidth

## 1 Introduction

Cloud computing is a product of the development and convergence of conventional computer technologies and network technologies, such as grid computing, distributed computing, parallel computing, utility computing, network storage technologies, virtualization, load balance etc. [1]. To ensure that the users can easily and quickly use resources through various kinds of terminals, Cloud computing provides available, convenient, and on-demand network access method to manage the various resources from cloud effectively and safely according to the demands of users. As the growing user demands, how to effectively manage resources and make it quickly available to users becomes the key technology of cloud computing needed to be addressed.

Virtualization technology is the foundation of cloud computing, it is a kind of infrastructure and the upper system and software to coupling separation technology, Virtualization technology through the upper service package to the virtual machine, and

manage resources through the deployment of virtual machines. Virtual machine deployment is to map the physical resource based on the virtual machine deployment request according to a reasonable allocation rules. The whole process is to seek optimal deployment of physical hosts under a multi-constraint optimization problem. Therefore, effective virtual machine deployment model and algorithm will be the key point of the efficient use of resources.

With in-depth studies on virtual machine deployment algorithms, the resource mapping of Virtual machine and physical host is developed from earlier one-on-one to later more-on-one relationship. Besides, Virtual machine deployment also changed from a single virtual machine mode to the virtual machine cluster. Virtual machine cluster refers to the communication needs and deployment restrictions of multiple virtual machine deployment limitation [2]. However, the deployment of the virtual machine cluster still faces many problems, such as high network communication consumption between the virtual machine, physical host resources waste problem and so on.

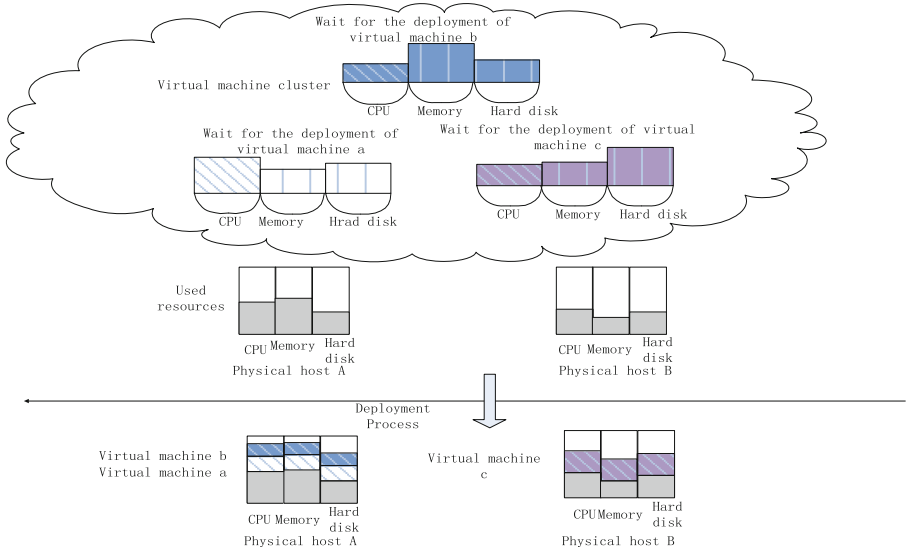
Many researchers have carried out extensive research on the virtual machine clusters deployment. For different application requirements of users, Paper [3] presents the sequence deployment strategy and balanced deployment strategy, however, these strategies only considering CPU resource constraints between virtual machines and physical host, wasting the physical host resources significantly. To solve the problem, paper [4] presents a resource matching strategy based on CPU and memory. Compared to paper [3], the strategy improves the usage of the physical host resources. However, it did not fully take into account the composition of virtual machines from a variety of resources, so this strategy cannot meet the needs of users in a variety of applications. To satisfy the various requirements of users, Paper [5] proposed a performance vector-based algorithm for virtual machine deployment. However, the deployment process only considers a single resource constraint of virtual machine and physical hosts. When more virtual machines exist, although the resource waste reduction in the rate of a single physical host, the waste of resources in the rate of the overall system does not reduce. Paper [6] presents a heuristic algorithm based on graph decomposition, which only considers the deployment of virtual machines on a single physical host. To cope with the problem, paper [7] proposed a decomposition algorithm. Throughout the deployment process, the overall rate decreased waste of resources. Paper [8] proposed a decomposition algorithm. The paper has associated physical resources as a deployment target, but only considers the communication bandwidth factor.

For the virtual machine cluster deployment issues, we proposed a MCSA based on constraint of resources and communication bandwidth. The algorithm firstly quantifies the virtual machines cluster resources and bandwidth, so that the virtual machines cluster can construct a weighted directed graph. As the weights of the nodes depict resources and the weight of each edge represents the value of the communication bandwidth. The double constrained optimization problem of resources and communication bandwidth can be translated to sub-chart graph partitioning problem. With minimum cut algorithm, the virtual machine is divided into small clusters from virtual cluster. For each virtual machines cluster, if its external communication is larger, the internal communication is smaller and vice versa. Then we calculate the cluster resource matching distance of the

virtual machine cluster and the physical host, which can be determined the approximate optimal solution of deployment issues.

## 2 Related Works

Virtual machine is composed of resources (CPU, memory, hard disk and other resources) which are required by the user. We can consider the virtual machine as entities consisted of all kinds of resources, in which physical host is serving as containers. Virtual machine deployment is to establish a resource mapping between virtual machines and physical host, then under the constraint of related resources, the virtual machine will looking for the best physical host to deploy. With the Deeping of the study, virtual machine deployment has developed from a single virtual machine deployment to the virtual machine clusters deployment, as shown in Fig. 1 [5].



**Fig. 1.** Virtual machine cluster deployment model

In the cloud computing environment, application providers usually deployed services in the physical hosting, and these services are generally made by the end user to a virtual machine. In order to efficiently provide service to users, Virtual machines need to collaborate with each other to jointly complete the user’s needs. Thus, a plurality of virtual machines with communication requirements and deployment restrictions constitute a virtual machine cluster. Figure 1 describes the virtual machine cluster deployment model. Multiple virtual machines constitute a virtual machine cluster, virtual machines and physical host are component with the CPU, memory, hard disk, and other resources. In standby mode, the physical host requires certain resources to run the initial state, removing the physical hosts in the standby state of the resource, by calculating the remaining available resources of the physical host resources available to meet the

physical host in the case of virtual machine resources required, will deploy a virtual machine to a physical host. Different deployment constraints, the results for the entire deployment will have greater impact.

### 3 Virtual Machine Cluster Deployment Algorithm

#### 3.1 Related Terms

Virtual machine cluster deployment description:  $n$  mutually between communication bandwidth demand virtual machine cluster,  $m$  need to deploy to the physical host.

In order to better describe the problem, we introduce the following symbols (Table 1).

**Table 1.** Basic terminology

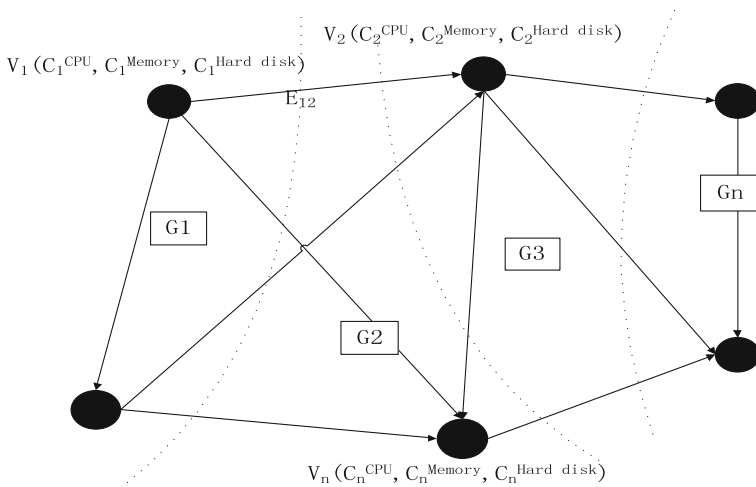
Vars	Description
$N$	The number of virtual machines
$M$	The number of physical hosts
$H_j$	The set of physical host to deploy the virtual machine
$C_i^{CPU}$	The CPU total capacity of the physical host $i$
$C_i^{Memory}$	The memory total capacity of the physical host $i$
$C_i^{Hard\ disk}$	The hard disk total capacity of the physical host $i$
$Q_j^{CPU}$	The CPU demand for virtual machine $j$
$Q_j^{Memory}$	The memory demand for virtual machine $j$
$Q_j^{Hard\ disk}$	The hard disk demand for virtual machine $j$
$P_{ij}^{CPU}$	The CPU match value of the physical host $i$ between virtual machine $j$
$P_{ij}^{Memory}$	The memory match value of the physical host $i$ between virtual machine $j$
$P_{ij}^{Hard\ disk}$	The hard disk match value of the physical host $i$ between virtual machine $j$
$E_{ij}$	the figure of the edge weights for virtual machine $i$ between virtual machine $j$
$D_{ij}$	The Resource matching vector for virtual machine $j$ between physical host $i$
$K_{ij}$	The desired physical host Weighted match vector for virtual machine $j$ between physical host $i$

#### 3.2 Virtual Machine Cluster Deployment Model

Virtual machine cluster for physical host deployment process works as follows: first, according to the communication bandwidth constraints segmented virtual machine cluster, getting a virtual machine cluster divided. Second, calculating resource matching between the virtual machine and physical host clusters, and searching for the best physical host

deployment. The deployment process of resources and communication bandwidth constraints need to be considered at the same time.

1. According to the communication bandwidth constraints, we use the minimum-cut algorithm [9] for virtual machine cluster segmentation. After segment, it forms the plurality of virtual machine cluster. The Segmentation strategies are as follows:
  - (1) Virtual machine is represented as a vertex graph.
  - (2) Communication bandwidth relationship between virtual machines is defined as the Edges.
  - (3) The virtual machine's resources (such as CPU, memory, hard disk) is expressed as the figure of peak value  $V_i(Q_j^{CPU}, Q_i^{Memory}, Q_j^{Hard\ disk})$ , the communication bandwidth between the virtual machine is expressed as the figure of the edge weights  $E_{ij}$ . Through the above quantitative, the problem of Virtual machine cluster can be converted into a problem of weighted undirected graph.
  - (4) The division process of virtual machine cluster is transformed into segmentation process of diagram. The segmentation of process is divided by the minimum-cut algorithm, so graph G can be divided into a plurality of sub-graph G1, G2... Gn. As shown in Fig. 2.



**Fig. 2.** Weighted undirected graph

2. Computing resource requirements of the virtual machine cluster, according to the physical host resources condition, seeking the best physical host through resource constraints.

The virtual machine cluster converted into a weighted undirected graph, we use the minimum-cut algorithm to divide weighted undirected graph, and get a number of weighted undirected graph after divided. The weighted undirected graph represents the virtual machine cluster. Next, Virtual machine cluster choose physical host to deploy.

In order to improve resource utilization, we should choose the optimal physical hosts to deploy. In this paper, we use the Euclidean distance cluster to represent the virtual machine and physical host resources matching degree.

The various resources matching:

$$P_{ij}^{\text{CPU}} = \sqrt{C_i^{\text{CPU}^2} - Q_j^{\text{CPU}^2}} \quad (1)$$

$$P_{ij}^{\text{Memory}} = \sqrt{C_i^{\text{Memory}^2} - Q_j^{\text{Memory}^2}} \quad (2)$$

$$P_{ij}^{\text{Hard disk}} = \sqrt{C_i^{\text{Hard disk}^2} - Q_j^{\text{Hard disk}^2}} \quad (3)$$

Calculating the virtual machine and the physical hosts a variety of cluster resources match, getting resources match vector  $D_{ij} = (P_{ij}^{\text{CPU}}, P_{ij}^{\text{Memory}}, P_{ij}^{\text{Hard disk}})$ .

In order to meet the needs of users for different resources, we weighted distance vector to represent user demand for resources. Weighted vector  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ . Finally, get the desired physical host Weighted match vector:

$$K_{ij} = \alpha * (P_{ij}^{\text{CPU}}, P_{ij}^{\text{Memory}}, P_{ij}^{\text{Hard disk}}) \quad (4)$$

Destination physical host Match:

$$L = \sum_{j=0}^m K_{ij} \quad (5)$$

Type L is the matching degree of the physical host between virtual machine clusters. With the L value reduced, the match degree of the physical host between virtual machine clusters is become well.

### 3.3 System Communication Bandwidth Utilization Rate

Communication bandwidth utilization rate (The communication bandwidth occupancy rate) R refers to the value of virtual machine cluster bandwidth utilization and the ratio of the communications bandwidth in the whole system.  $B_w(V_i, V_j)$  represents the communication bandwidth between two virtual machines, T represents the virtual machine cluster deployment environment. On the process of virtual machine deployment to host, some virtual opportunities deployed on the same host, for the external communication bandwidth of the virtual machine transformed the internal communication of the host. The communication of the whole system will not result in a greater impact. Therefore, when the virtual machine is deployed to the different physical host, T is 1, when the virtual machine is deployed in the same physical host, T is 0. And  $B_n$  represents the total bandwidth. So Communication bandwidth utilization rate is expressed as:

$$R = \sum_{i=1}^n B_w(V_i, V_j) * T / B_n \quad (6)$$

### 3.4 The Analysis of System Resource Waste Rate

Because of different deployment strategies, there are large differences in Virtual Machine Deployment Results. For example, when the distribution of resources on the physical host is uneven, it is like to cause the entire system to waste resources. So, in this article, we calculate the physical host resources waste rate to reduce the waste of the resources of the system. Wastage rate  $Waste_p$  refers to the average ratio value between different resources and the whole physical server, as shown in Equation.

$$Waste_{p_j}^{CPU} = (p_j^{cpu} - Q_j^{CPU})/p_j^{cpu} \quad (7)$$

$$Waste_{p_j}^{Memory} = (p_j^{Memory} - Q_j^{Memory})/p_j^{Memory} \quad (8)$$

$$Waste_{p_j}^{Hard disk} = (p_j^{Hard disk} - Q_j^{Hard disk})/p_j^{Hard disk} \quad (9)$$

Optimized resource wastage rate can be expressed as:

$$\min Waste_p = \sum_{j=0}^m (Waste_{p_j}^{CPU} + Waste_{p_j}^{Memory} + Waste_{p_j}^{Hard disk})/3 \quad (10)$$

### 3.5 Virtual Machine Cluster Deployment Algorithm Process

In this paper, algorithm is based on double constraints of the resources and communication bandwidth, and we quantify the resources and bandwidth to form a weighted undirected graph, where the vertex weights graph represents the resources, right side of the figure represents the value of the communication bandwidth, the double constraints of resources and communication bandwidth optimization problem can be transform into a graph of graph partition problems, we use the minimum cut algorithm to break up the Weighted undirected graph. Next, calculate the approximate solution of the problem.

- (1) Initialize the data center, randomly generated virtual machine and the physical host. Get resource requirements of the virtual machine cluster  $R_{VM}$ , communication bandwidth between the virtual machine  $W_{VM}$ , the list of hosts' available resources of all hosts  $H_{PM}$ .
- (2) Modeling for virtual machine cluster deployment problem, quantify the virtual machine cluster, get weighted undirected graph.
- (3) Converted the virtual machine cluster to weighted undirected graph and use the minimum-cut algorithm to divide weighted undirected graph.
- (4) After segmentation, calculate the resources matching distance of virtual machine cluster and physical host calculation, if physical host can be deployed it, we will build the virtual cluster to deploy on physical host. If there is no, then jump step (3).
- (5) Cycle all virtual machines cluster list, until all the virtual machines clusters deployed over.

The deployment process is shown in Fig. 3 below:

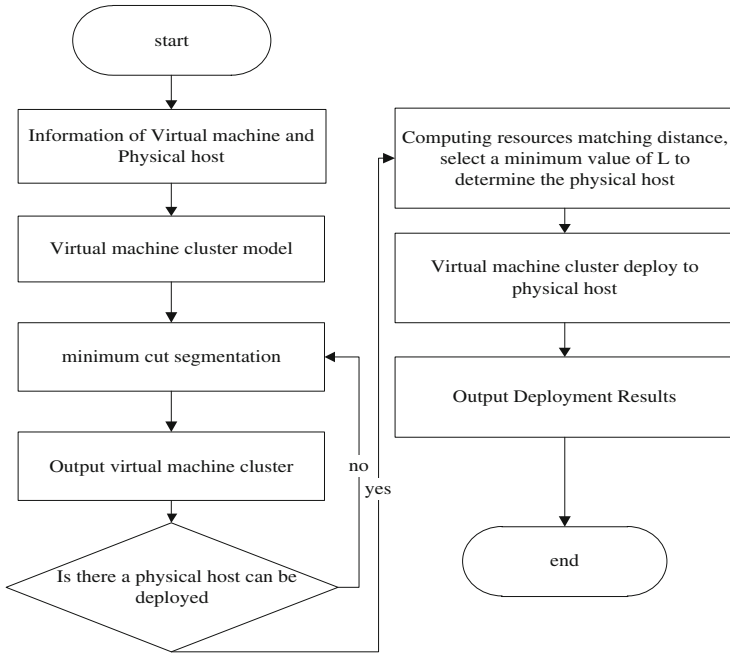


Fig. 3. Virtual machine cluster deployment algorithm flow chart

The Algorithm is described as follows:

**For** every VM

    Get VM resource requirements data list  $R_{VM}$ , the communication bandwidth of VM array  $W_{VM}$

**For** every PM

    Get PM performance data list  $H_{PM}$

**Label 1 : For**  $W_{VM}, R_{VM}$

    Min-Cut algorithm ()

    VM clusters  $R_{VM}$  Divided into  $R_{VMi}, R_{VMj}$

    Array  $W_{VM}$  Divided into  $W_{VMi}$  and  $W_{VMj}$

**For**  $R_{VMi}$  or  $R_{VMj}$  resource requirements data matching to  $H_{PM}$  performance data

**If** match success, get current matching degree list L

            Get the  $L_{min}$

            Then get the best match PM, and return deployment result

**Else** match failure

            Goto label 1

**Return** all deployment result



## 4 Simulation and Analysis

In this paper, we conduct simulations based on the Cloudsim 3.0 [10] with the operation of windows 7 64-bit. The JDK version adopted in the paper is jdk1.6.0\_43. We compare greedy algorithm, a single resource constraint algorithm and article of virtual machine cluster allocation algorithm to analyst is System resources waste rate and Communication bandwidth occupancy rate. Simulation results validate that the algorithm has a good performance compared to other algorithms.

### 4.1 Simulation Platform

For the simulation platform, according to this paper, we have expanded our simulation platform by recompiling CloudSim3.0.3, and the writing simulation programs. First, we initialize a data center, and each data center contains a number of physical hosts. In the data center, we use using a random way to produce physical resources and virtual machine hosts. At the same time, by expanding the classes, Datacenter, Host, virtual machine and DataCenterBoker, we realize the underlying physical and virtual machine simulation. The experiment procedure of physical machines and virtual machine strategies are as follows:

(1) Virtual machine

Using random strategy, generate virtual machine allocation request queue, in which the CPU is generated randomly from 1 to 6 nuclear, memory and hard disk are also randomly generated. For each generation of virtual machine memory, the quantity is 512 M integer times and hard disk is the integer times of 16 G.

(2) Physical host

Custom Datacenter Characteristics class, generate the corresponding Datacenter and physical Host. Including CPU, memory, hard drive 10 integer times randomly generated by the virtual machine.

(3) The communication bandwidth of virtual machine.

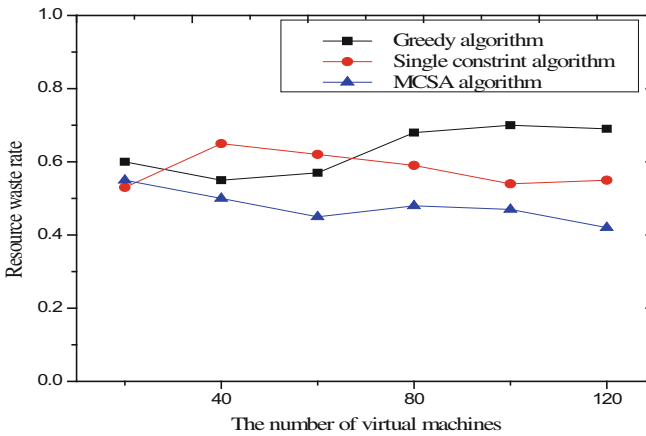
Using randomly generated strategy, generate virtual machine communication bandwidth matrix between 0–9.

### 4.2 Results Analysis

1. System resources waster rate

System resources waste rate refers to the average resources waste rate of the physical hosts deployed with virtual machines. It is to note that in the paper, we only consider the CPU, memory and hard disk. This can measure the system resource utilization. Figure 4 depicts the physical host resources waste rate differences between the various algorithms. The Fig. 4 shows that with the increasing of virtual machine requests, the algorithm proposed in this paper can gradually reduce system resource waste rate and tends to be stable. For the three kinds of algorithms, the resource waste rate of MCSA algorithm is the lowest, followed by the single resource constraints algorithm, greedy algorithm is the worst. As we can see from the figure,

MCSA algorithm has good performance lies in that when allocating resources, we taken the approximation degree of virtual machine cluster between physical hosts into consideration. Specifically, if the degree is closer, it means more balanced use of resources after the distribution of the physical host, the greater the variety of resources available extent and the smaller the rate of physical hosts waste of resources. As the greedy algorithm does not adopt any optimization mechanism in resources allocation, it has the highest waste rate.



**Fig. 4.** system resources waste rate

## 2. System bandwidth occupancy rate

Communication bandwidth occupancy rate refers to the ratio between the communication bandwidth of each virtual machine cluster and communication bandwidth needed for the whole system in the physical host. It represents communication bandwidth occupying degree of the whole system when running a virtual machine cluster. As it can be seen from the Fig. 5, with the increasing number of virtual machines, MCSA algorithm can keep the communication bandwidth occupancy rate at a low level. The reason lies in that it divides the virtual machine cluster with into several virtual machines cluster with a minimum cut algorithm which has lower communication and bandwidth demands. Meanwhile the virtual machine cluster with larger communications bandwidth demands redeployed on the same physical host. The single constraint algorithms occupy larger communication bandwidth as it only considers the resources of the virtual machine between cluster and physical host. The greedy algorithm did not consider any allocation optimization. It has the largest communication bandwidth occupied.

To sum up, in the situation where the virtual machine in the cluster requires frequent communication, the proposed virtual machine cluster deployment algorithm in the paper can keep a low system resource waste rate, stays small system communication bandwidth, and achieves high network utilization.

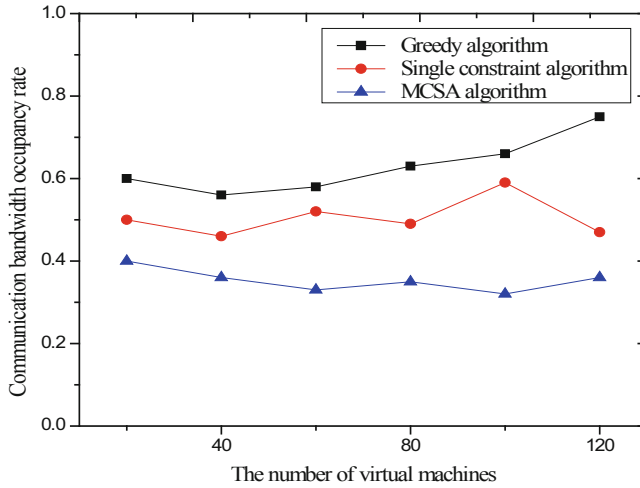


Fig. 5. System bandwidth occupancy rate

## 5 Conclusions

In this paper, to cope with the Virtual machine cluster deployment issues in cloud computing platform, we translate the virtual machine cluster deployment into optimization problems under multiple constraints. We presented a MCSA algorithm based on the double constraints of virtual machine resources and communication bandwidth. The algorithm firstly quantifies the resources and the communication bandwidth in the virtual machine cluster and separates the virtual machine cluster by minimum cut algorithm of graph theory. Then based on the segmentation of virtual machine cluster, the algorithm can effectively select the target physical host. The simulation results validate that it can reduce the resource waste rate and the system communication bandwidth utilization rate significantly. For further research, we aim to explore the combination of virtual machine energy consumption and resource equilibrium problems.

**Acknowledgment.** The authors would like to thank the anonymous reviewers for their helpful and constructive comments. This work is supported by education reform Item of Hunan Normal University (Grant [2014]75). Program for Excellent Talents in Hunan Normal University (Grant no.ET61008).

## References

1. Tu, H.-K., Zhou, H., Lin, R.-H.: Design and implementation of enhanced parallel computing framework system in cloud. *New Ind.* **12**, 33–40 (2012)
2. Wang, G., Ma, Z., Sun, L.: Deployment of virtual machines with clustering method based on frame load awareness. *J. Comput. Appl.* **33**(5), 1271–1275 (2013)
3. Yuan, J.: *The Research on Multi-VM Fast Deployment Mechanism*. Huazhong University of Science & Technology, Wuhan (2008)

4. Cao, W., He, J., Sun, Z.: Research on Mechanism of Deployment Virtual Machine in Mode of IaaS. *Comput. Technol. Dev.* **10**, 105–108 (2012)
5. Yang, X., Ma, Z., Sun, L.: Performance Vector-based algorithm for virtual machine deployment in infrastructure clouds. *J. Comput. Appl.* **32**(1), 16–19 (2012)
6. Meng, X., Pappas, V., Li, Z.: Improving the scalability of data center networks with traffic-aware virtual machine placement. *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9. IEEE (2010)
7. Nguyen Van, H., Dang Tran, F., Menadue, J.M.: Autonomic virtual resource management for service hosting platforms. In: *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*, pp. 1–8. IEEE Computer Society (2009)
8. Zhu, Y., Ammar, M.: Algorithms for assigning substrate network resources to virtual network components. In: *INFOCOM 2006, 25th IEEE International Conference on Computer Communications, Proceedings*, pp. 1–12. IEEE (2006)
9. Zhang, J., Li, B.: Research of image segmentation based on graph theory and minimum cut set algorithm. *Laser Technol.* **29**(6), 863–866 (2014)
10. Calheiros, R.N.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw. Pract. Exper.* **41**(1), 23–50 (2011)