ADAPTED BIN PACKING ALGORITHM FOR VIRTUALS MACHINES PLACEMENT INTO DATACENTERS

Fréjus A. R. Gbaguidi[1, 3], Selma Boumerdassi[1, 2], and Eugène C. Ezin[3]

[1]Conservatoire National des Arts et Métiers / CEDRIC, Paris, France
[2]INRIA Hipercom, Paris, FRANCE
[3]IMSP /Universitéd'AbomeyCalavi, Benin

ABSTRACT

The placement of virtual machines is a permanent routine that determines both performance and energy efficiency within Datacenters. Unfortunately, it is a task whose complexity is fully supported by the common sense of the system administrators who must try different scenarios in order to detect the one that best satisfies the constraints imposed by the environment. Bin packing techniques have been used to address similar issues in other areas such as transportation and mass distribution. We try to apply these methods to the problem of placing virtual machines on the physical servers within Datacenters. Our aim is to evaluate the efficiency of this technique at the optimum distribution of the VM while using the minimum number of physical machines and consequently reduce the amount of energy required for their power supply. The results obtained in comparison with the so called brute force method makes it possible to conclude that the Bin packing techniques could help possible to rationalize the use of the physical resources allocated to the operation of the applications in the Datacenters while preserving the SLA imposed by the clients

KEYWORDS

DataCenter, Energy consumption, Bin Packing, Cloud Computing, Datacenters.

1. INTRODUCTION

Virtualization technologies and cloud computing have emerged particularly during the last five years thanks to the widespread needs of ever faster, more sophisticated and always easier technologies. The formerly known enterprise self-computing disappears gradually in favor of outsourcing IT services and on-demand services consumption that both boost business and improve competitiveness in more complex environments. Early, the efforts of industrial and ICT professional shave turned to the improvement of basic technologies of Cloud Computing and first, virtualization. Considered as the forerunner of the modern computing, the possibility of subdividing a physical machine into multiple virtual servers as shown on fig 1 has generated among researchers appetites still not satisfied. Since that hypervisors have been made available, all manufacturers in the computing world migrated their platform in this new perspective, favoring the growth of many datacenters in which the customized services at an optimal price are now possible. The proliferation of data centers around the world, however, reveals the ever increasing needs of material and energy resources for their operation given the demands of the nonstop" generated outsourcing needs and the exponential growth of traffic volumes of online services.

Natarajan Meghanathan et al. (Eds) : ICCSEA, WiMoA, SPPR, GridCom, CSIA - 2017
pp. 69– 80, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71106

The question, that is the more worrying is that the energy needs of Datacenters now rival the most energy-intensive industries such as aerospace, automotive, to name a few. Many studies have been initiated to try to reduce the magnitude of the problem then generated. First, the hardware manufacturers have invested in the development of components with high electrical efficiency. Then, other software adjustment techniques of the server operating system helped to maximize energy savings. However, in order to respect the constraints of availability and cloud computing services quality, engineers require to size the virtual servers with comfortable margins for avoiding congestions and saturations of processors and memory resources. Energy losses caused by that operation of resources oversized become a new nightmare to which many studies trying to answer. Several algorithms were then developed to optimize virtual machine allocation plan on physical servers. Overall, these algorithms are based on solving a bin packing problem in multiple dimensions with different variants. Two main methods of solving these types of NP-hard problems clash.

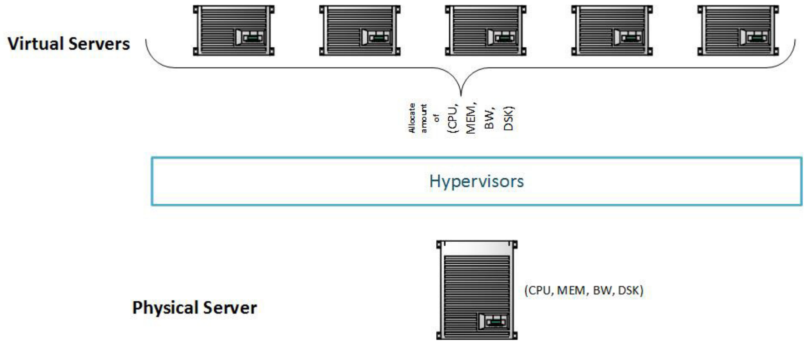


Figure 1. Virtualization classical scheme

On the one hand the so-called "bruteforce" causing very high computation time and incompatible with the process of rapid decision making in the field of cloud computing and on the other hand the heuristic method that combines some logic to find the most realistic solution within an acceptable time. However, the placement problem of virtual machines in the datacenter must take into consideration certains others conditions such as incompatibilities, compliance of combined value memory-processors to a minimum and respect of Service Level Agreements (SLA) constraints. Here we propose a new algorithm that is based on heuristics and produces an occupancy plan of virtuals machines in the datacenter by minimizing the overall processors amount and therefore improve energy efficiency within datacenters. This techniques also help to reduce the decision time for datacenters engineers comparing to manual or brute-force methods.

The contribution of our work is about, on the one hand, the use of the Bin packing algorithm for the design of initial allocation plan of the virtual machines and other hand the introduction of the compatibility constraint regarding to the high availability and data security architectures implemented within Datacenters. The rest of the article is organized as follows: in section 1 we present the state of the art on VM placement techniques, and then describe the Bin Packing tools used in the section 2. We will present our placement approach in Section 3, and the presentation of simulation results in Section 4. Finally we'll conclude the paper.

2. RELATED WORKS

The virtual machine placement algorithms within Datacenters are purchased different goals. Based on several survey [9],[12], [7], [10] [6] the most common one is the reduction of the amount of physical resources deployed and thus optimizing the overall energy consumption. The techniques developed are largely based on heuristics or meta-heuristic of the well-known Bin

Packing problem solving but other techniques such as vector and matrix calculations or methods of artificial intelligence as ant colony systems or Bayesian systems are also used. Many papers like [3] presents a variant of Bin Packing known as the snapsack problem developed to solve de placement problem.[13], [4] proposes an algorithm based on ant colony systems to develop a distribution function of the triplets (cpu, mem,bw) corresponding to the needs of the triplets VM (CPU,MEM, BW) corresponding to available servers. In this article, the utilization rate criteria different physical resources are taken into account in respect of decision making. Similarly, others techniques are proposed by [1], [11] and are based on an iterative approach of VM placement in purpose of cloud service providing. In a more improved version closer to the needs of real data centers, [2] conduct a prior classification of servers resource consumption profile before deciding their distribution on physical servers. This work helps distribute virtual servers to limit sudden later resources congestion on physical servers. The various techniques presented above using various mathematical tools. Techniques based on binpacking problems are most common in the literature and at the same time those that produce the best results. However the classification proposed by [2] does not necessarily correspond to the stresses that are usually found in data centers. We will try to add a manual classification of the servers according to their compatibility.

3. BIN PACKING PROBLEM

Given a set of objects of rectangular shapes of any known size and given a larger rectangular form bin of known dimensions, the bin-packing problem (BP) is to determine the minimum number of bins needed for store without overlap all of these objects (objects do not extend beyond the bins and do not overlap) [8], [5], [6]. More formally, the binpacking problem(BP) is defined as follows: given a set of n rectangular objects A = {a1,…, an} and an unlimited number of identical rectangles (the bins) of larger dimensions than those of the objects, the problem is to determine the minimum number of bins used to store all objects without overlapping. The problem of bin packing can be approached according to different aspects that we expose below

3.1. Different cases of bin packing problem

Depending on the shape of the treated objects, the binpacking problem can occur in one, two or multi-dimension

- 1BP (one dimension): Is to minimize the number of one dimensional containers (bins) needed to store a list of items characterized by their length

- 2BP (two dimensions): Is a generalization of 1BP. This is to minimize the number of large identical rectangles(bins) needed to store a rectangular shape item list. The items must be arranged in such a way that the sides of the rectangles are parallel to those of the bins

B. Different types of complexity

The problem of storage can be complex depending on the sometimes irregular forms of items or other types of constraints imposed by the problem

- objects of homogeneous or heterogenous forms;

- objects of uniform or different sizes;

- deformable or non-deformable objects

- the number of dimensions of the problem;
- have a single bin (decision problem or maximization problem);
- seek to minimize the number of bins to use;
- seek to minimize the surface or the volume of objects to be placed;
- linear constraints between objects;
- constraints on the order in which objects must be removed from the bin;
- orientation constraints of an object;
- weight constraints (eg the weight of a full bin can not exceed a certain limit)
- Investment constraints, some very heavy objects should be placed at the bottom, other vulnerable must be placed above;
- orientation: objects can be fixed orientation (one speaks of the oriented case) or they can be rotated 90 degrees(the undirected case)
- The guillotine constraint: If it is imposed, it must have the possibility to return the items stored by end cuts parallel end to the dimensions of bins;
- Incompatibility : Some objects can not coexist in the same bin;
- etc.

C. Different solutions

Solving bin packing problems can be very complex and expensive in terms of computing time when it comes to exact solutions. There are fortunately heuristic solutions to approach the problems by providing reliable and very fast results

- Approximate methods: no optimal solution but rapid resolution. Choice of algorithm depending on the problem
- Heuristics algorithms 1 BP
    - Next Fit: assignment of the current object has in the current box if wishes. Otherwise, closing the current box, opening a new one
    - First Fit: allocation of current object has the available box
    - Best Fit: assignment of the current object in the best box
    - Worst Fit
    - Any Fit
- Heuristics algorithms 2 BP

  - In one phase directly arrange the objects in the BIN

  - In two phases: preliminary resolution of strip-packing problem (storage of all objects in a bin without height limit) starting by order of Decreasing height before storing in the bin; and apply one of the above methods. We'll talk about NFD, FFD or BFD

- Metaheuristics

- Several others methods like taboo Search or LowerBounds

- Exact methods: due to the NP-hard features of Binpacking problem, they resolution time by exact methods is too long and expensive. Facing the lack of a universal algorithm, linear programming is often use to solve problems

4. VM PLACEMENT PROBLEM

4.1 Stating the VM placement problem

Distribute virtual machines on physical servers may seem at first sight a simple division operation of the physical server resource capacity by the same amount of those resources required by virtual machines. Number of data center administrator is exercised manually to resolve the issue by trying to the best of their ingenuity to efficiently allocate resources to VMs. However, with the big size datacenters, it is essential to find tools on the one hand to automate this task and streamline the allocation of resources on other hand. To do this, the distribution of resources is generally reduced to a bin packing problem in which we compute the quantities required for each n resources of the objects to be stored for this dimension while the size of said resource on the physical server is the storage bin. The dimensions N could be represented by trying upto solve the problem in one time for all the n dimensions but we go here to a resolution per dimension. This means finding, based on CPU demands of virtual servers, the number of physical servers required to house them. The standard form of mathematical expression of the problem is as follows.

P is the set of physical servers,

V all virtual machines to distribute,

Rij the amount of resource i on Vj

Tij , the total size of the resource i on Pj

This is to find an allocation matrix M with terms Xij = 1if VM i is placed on the PM j and equal 0 otherwise.

Alloc =Min(P) and Xij = 1

This matrix is made with the following conditions:

1) The amount of the resource for all VMs placed on a given server can not exceed the total size of the resource on said server

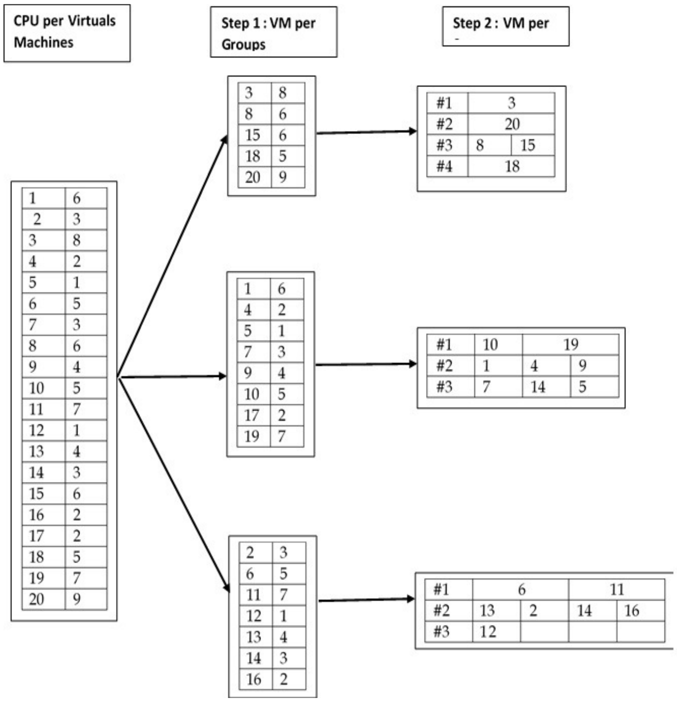
2) All VM have to be placed. Thus, the sum of Xij = 1whatever i

3) Allocate the VM on the optimal number of PM. This means finding the minimization function that gives the smallest possible result from a set of solutions.

We add to this classic problem a clause of in compatibility between certains VM. Specifically, it is to solve the technical constraints imposing a distribution of redundant servers in a cluster on different physical machines so that in case of failure of one, continuity of service is automatically provided by the other. To do that we will implement a pre treatment which is to classify the servers depending on their compatibility.

4.2 Our VM placement approach

The classic problems of Bin Packing typically focused onBin of identical size. We consider, for simplicity, that all physical servers have the same technical characteristics.



1	6
2	3
3	8
4	2
5	1
6	5
7	3
8	6
9	4
10	5
11	7
12	1
13	4
14	3
15	6
16	2
17	2
18	5
19	7
20	9

3	8
8	6
15	6
18	5
20	9

#1	3	
#2	20	
#3	8	15
#4	18	

1	6
4	2
5	1
7	3
9	4
10	5
17	2
19	7

#1	10	19	
#2	1	4	9
#3	7	14	5

2	3
6	5
11	7
12	1
13	4
14	3
16	2

#1	6		11	
#2	13	2	14	16
#3	12			

Figure 3. Our Placement steps

We divide the problem into two parts which are:

- Classification by compatibility group virtual servers This is handled semi-automatically. An operator must manually register the virtual machines within a given group. It will be possible to automate this task only through a form to fill out to keep each VM on it compatible group. Then, an algorithm can be implemented for compatibility checks. However, the decision to separate virtual servers can be efficiently taken by an engineer with knowledge of the security constraints, and continuity of service required in Datacenter environments to match the SLA.

- Determining the number of physical servers required per group and consolidation of virtual servers within physical server (see figure 3) Our approach at this level is based on an adaptation of the method of the FirstFit Decreasing (FFD) for allocation of virtual machines on physical servers. Heuristic FFD is resolved into two phases namely: The first phase is to store all resource values required by the VM in a descending order. Then the First Fit algorithm is applies to successively add VM to physical servers until the limit allowed is reached ina second phase. Thus, VM are stored on the servers to the overall allocation of the list of hosting queries

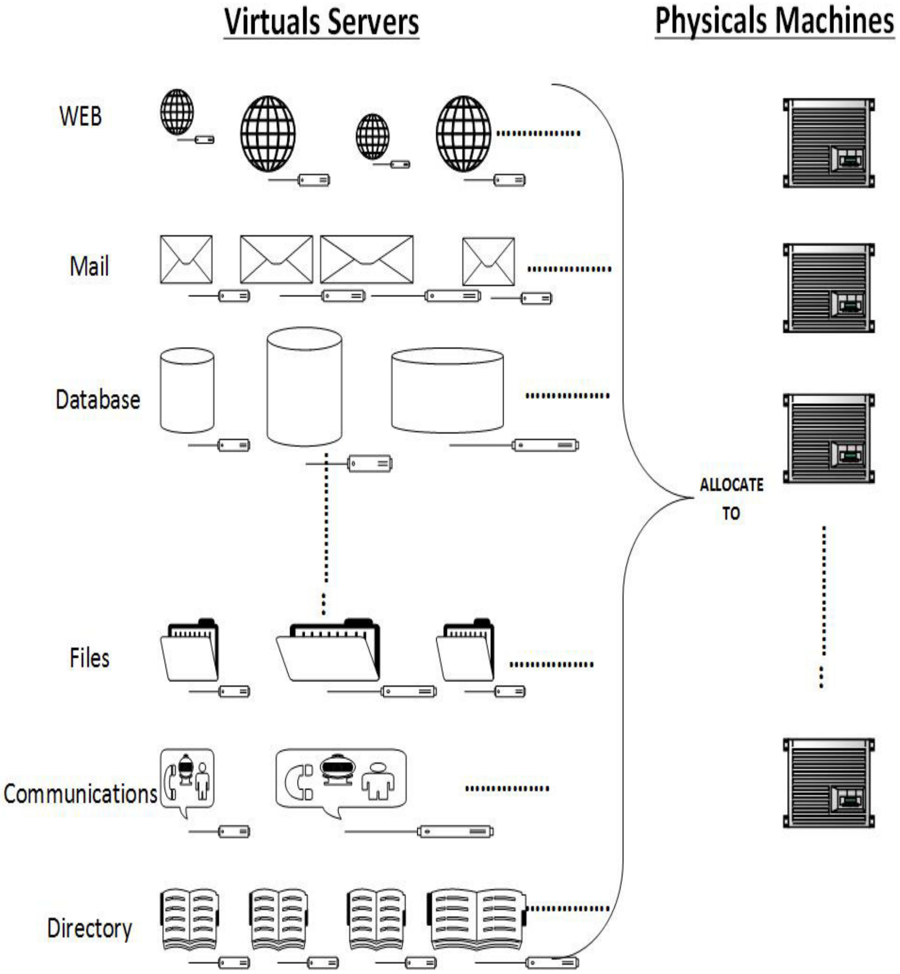
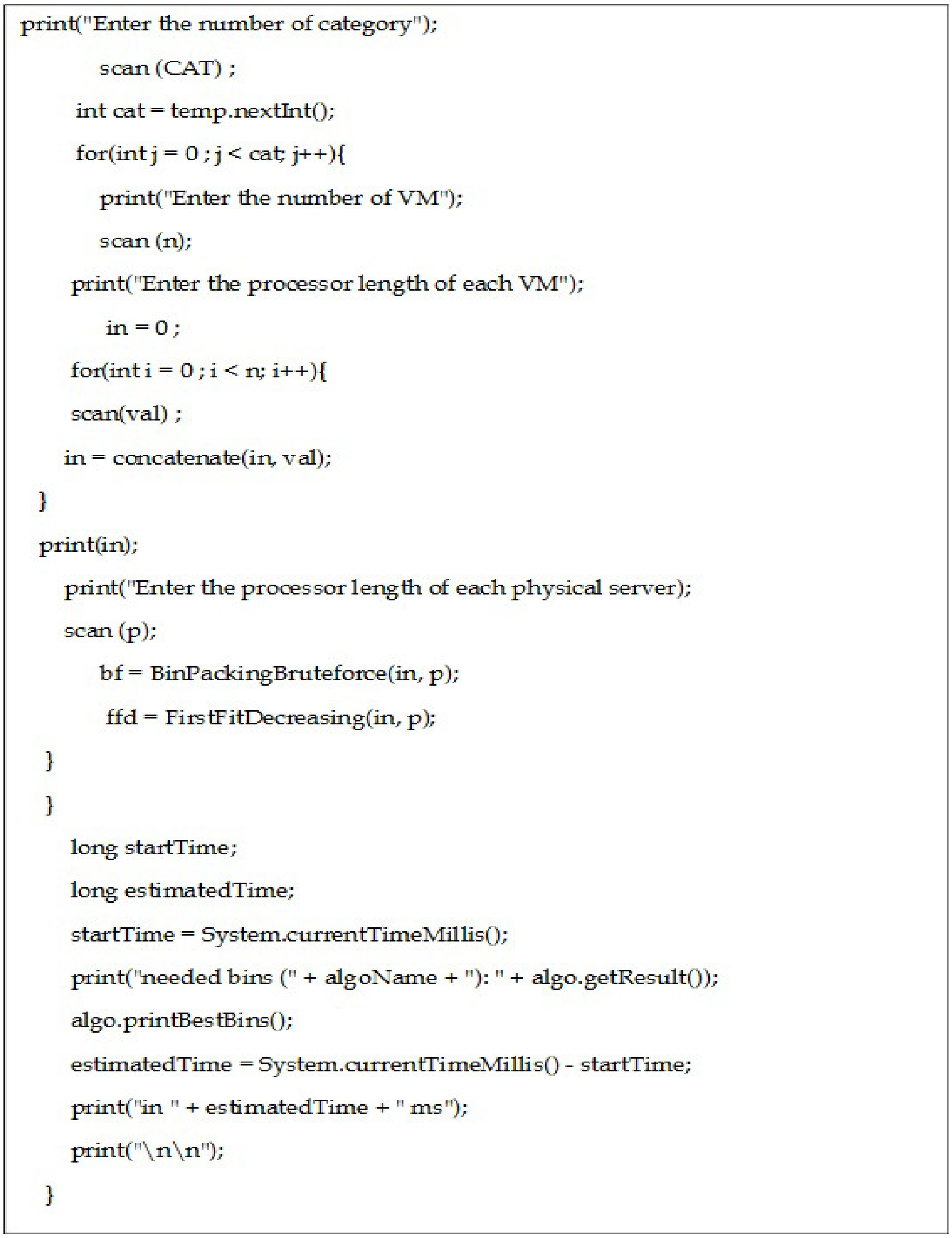


Figure 2. VM Placement Problem

The algorithms of the First Fit Decreasing and Best Fit Decreasing strongly compete in solving similar problems. While in the BFD an optimality criterion is added for the selection of Bins, FFD proceeded to the allocation bin after bin only. Assuming all Bin are identical in size, this additional step implemented in the BFD does not add any efficiency. It's then useful to choose the FFD to save in terms of processing time and thus the operation performance. Since datacenters requires real time operating, the resulting processing time saving fully justified the choice of FFD.

4.3 Placement algorithm

The algorithm is divide into severals procedure including Brute Force, First Fit Decresing, which compute each one the VM placement. The main program is presented below



```
print("Enter the number of category");
          scan (CAT) ;
       int cat = temp.nextInt();
       for(int j = 0 ; j < cat; j++){
          print("Enter the number of VM");
          scan (n);
      print("Enter the processor length of each VM");
           in = 0 ;
      for(int i = 0 ; i < n; i++){
      scan(val) ;
     in = concatenate(in, val);
  }
  print(in);
     print("Enter the processor length of each physical server);
     scan (p);
         bf = BinPackingBruteforce(in, p);
          ffd = FirstFitDecreasing(in, p);
   }
   }
      long startTime;
      long estimatedTime;
      startTime = System.currentTimeMillis();
      print("needed bins (" + algoName + "): " + algo.getResult());
      algo.printBestBins();
      estimatedTime = System.currentTimeMillis() - startTime;
      print("in " + estimatedTime + " ms");
      print("\n\n");
   }
```

Figure 4. Placement Algorithm

5. EXPERIMENTS AND RESULTS

5.1 Experiments framework

We try to demonstrate through simulations the performance of our approach to solving virtual machine placement problem by comparing an implementation of FFD with a brute-force method. The problem of Bin packing is NP-Hard, then the resolution by the brute force method consists on going to each occurrence of virtual machine and then determine the ability to allocate or not to the physical machines.

Our study framework is one of the institutional Datacenters in Benin in which the several publics administrations regularly requests hosting their application. It hosts at ually a hundred of servers but we will limited our study of a dozen HP Blade BL 350 with identical specifications in purpose to meet the constraints of the Bin Packing problems. Servers characteristics are as follows:

- Processor: Dual processor with six cores at 2,4 GHz eachper server
- RAM memory :128 Gbytes
- Hard disk : this resources are use over the SAN systemwith over 100 TBytes of capacity
- Network Controller: 10 Gbps

To facilitate data processing into our program, we voluntarily libellons the amounts of such resources unless decimals parts. Virtual machines on it is requested through a hosting application in which customers can complete their needs among other, amounts of CPU, Memory capacity, HDD capacity and number and speed of network controllers. Here again, we limit ourselves to twenty virtual machines requests. The summary of the resource requests of our experiments can be as follows:

Table 1. SERVERS LIST BY GROUP.

N°	**CPU Request**	**Group**
1	6	1
2	3	1
3	8	1
4	2	1
5	1	1
6	5	1
7	3	1
8	6	2
9	4	2
10	5	2
11	7	2
12	1	2
13	4	2
14	3	3
15	6	3
16	2	3
17	2	3
18	5	3
19	7	3
20	9	3

The machines are virtualized using VMware hypervisorESXi 5.5. The aim of our simulation is to compare the number of currently used server in the datacenter using a manual method and Brute force with that obtained using our adapted version of First Fit Decreasing algorithm. The simulation is to enter into the program developed for this purpose, the needs of users of Datacenter and the available quantity of the resource at the server. The simulation output is to determine for each method, the number of necessary physical server as well as the processing time.

5.2 Results

In the table II, we presents for each VM packing method the total duration of the process and thee number of physical Machine (PM) required to host the groups of VM. For manual technique, the duration is marked as Non Available since it depend on the ability of each Datacenters engineers.

Table 2. ALGORITHMS PERFORMANCE

Methods	**Group 1**		**Group 2**		**Group 3**	
	PM	Duration	PM	Duration	PM	Duration
Manual	4	NA	4	NA	4	NA
Brute Force	4	0	3	875	3	31
FFD	4	0	3	0	3	0

Considering the results, we clearly see that the determination of the initial VM placement scheme is optimal by the method of Bin Packing. The evaluation also shows that the calculation time by the method of Brute-force is very high compared to the FFD at a very high factor. The preliminary classification of machines by compatibility group is also an important feature of our study as this reality is often ignored in many similar work even though it is a basic rule of VM placement within reals Datacenters Remembering that our sample is small for simplicity, it is useful to note that the allocation scheme determination duration can be several times higher when considering a large size Datacenters (hundreds or thousands of servers). The use of FFD based technique is essential to mitigate that weakness and respond effectively to the flow of hosting applications requests in the datacenter. Determining an optimal placement scheme of virtual machines in a data center represents only the beginning of the placement spots since the quantities of resources requested initially are very different to that consumed after the production start leading huge loss of resources. These under consumption result in a waste of electrical power because once setting up, VM consume nearly the amount of energy that they need when they are fully loaded. To overcome this, a periodic reallocation of virtual machines on physical servers is necessary. However, even if the methods of Bin Packing can always be helpful, other algorithms based on appropriate criteria are further necessary.

6. CONCLUSION

The optimization of energy consumption within the datacenter is an area where research is so far very fruitful. The key lies in virtualization, technical development to improve the electrical efficiency of deployed servers and better allocation of virtual servers. With constantly growing needs of cloud computing, the placement problem of virtual machines on physical servers deployed in data centers has become central. It is divided into two parts which are: i) the initial placement scheme and ii) the optimal reallocation based on the real needs of users over time. Our goal in this work was to propose an approach to the issue of initial placement through methods of Bin Packing. Thanks to results produced in other areas like logistics, algorithms based on bin packing carry promises for improving resource consumption on servers within datacenters. In fact, our tests on a sample taken from a real environment show compared to other traditional methods of VM placement on servers that the heuristic of the First Fit Decreasing reduces so sensitive waste of resources on physical servers. Added to this, the optimization of the time required for determining the proper placement scheme is also highlighted. This reduced the time for Datacenters administrators in intelligence decision namely for customer VM hosting since cloud computing technologies requires more responsiveness and more efficiency for better satisfaction. The results, however, are just an initial basis for the resolution of energy efficiency issues in the datacenters. The increasing size of these, the variety of constraints related to technology and customer requirements involve the development and validation in varied environments of other techniques in the objectives of dynamic management of VM allocation under production. Existing virtual machine migration methods must be improved by combining existing solutions with the Bin Packing tools.

REFERENCES

[1] MB Arya and Ajay Basil Varghese. A combined bin packing vm allocation and minimum loaded vm migration approach for load balancing in iaas cloud datacenters.

[2] Norman Bobroff, Andrzej Kochut, and Kirk Beaty. Dynamic placement of virtual machines for managing sla violations. In Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on, pages 119–128. IEEE, 2007.

[3] Ricardo SteghCamati, AlcidesCalsavara, and Luiz Lima Jr. Solving the virtual machine placement problem as a multiple multidimensional knapsack problem. ICN 2014, page 264, 2014.

[4] Eugen Feller, Louis Rilling, and Christine Morin. Energy-aware ant colony based workload placement in clouds. In Proceedings of the2011 IEEE/ACM 12th International Conference on Grid Computing ,pages 26–33. IEEE Computer Society, 2011.

[5] Richard E Korf. A new algorithm for optimal bin packing. In AAAI/IAAI, pages 731–736, 2002.

[6] Andrea Lodi, Silvano Martello, and Michele Monaci. Two-dimensional packing problems: A survey. European journal of operational research,141(2):241–252, 2002.

[7] Fabio Lopez-Pires and Benjamin Baran. Virtual machine placement literature review. arXiv preprint arXiv:1506.01509, 2015.

[8] EG Co man Jr, MR Garey, and DS Johnson. Approximation algorithms for bin packing: A survey. Approximation Algorithms for NP-Hard Problems, pages 46–93, 1996.

[9] ZoltánÁdám Mann. Allocation of virtual machines in cloud datacenters—a survey of problem models and optimization algorithms. ACM Computing Surveys (CSUR), 48(1):11, 2015.

[10] Kevin Mills, James Filliben, and Christopher Dabrowski. Comparingvm-placement algorithms for on-demand clouds. In Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on, pages 91–98. IEEE, 2011.

[11] Weijia Song, Zhen Xiao, Qi Chen, and Haipeng Luo. Adaptive resource provisioning for the cloud using online bin packing. Computers, IEEE Transactions on, 63(11):2647–2660, 2014.

[12] B Benita Jacinth Suseela. Survey on vm placement algorithms. InternationalJournal of Engineering Trends and Technology (IJETT), 6(7):349–352, 2013.

[13] Gaochao Xu, Yan Ding, Jia Zhao, Liang Hu, and Xiaodong Fu. A nove lartificial bee colony approach of live virtual machine migration policy using bayes theorem. The Scientific World Journal, 2013, 2013.

AUTHORS

Fréjus A. R. GBAGUIDI is a PhD student in Computer science at Cnam Paris and Abomey-Calavi University Benin (IMSP-UAC). With a Master Degree in Network and Information System, He's a Datacenter and network specialist with many skills in severals fields including Information System Security and Audit.

Selma BOUMERDASSI is Associate Professor, at Conservatoire National des Arts et Métiers (CNAM) Paris in security and localization in ad-hoc networks and also Researcher at Institut National de Recherche en Informatiqueet Applications (INRIA-France)

Eugène C. EZIN is a Lecturer of the universities of CAMES, Director of Institut de Formation et de Recherche en Informatique at Abomey Calavi University. His research fields include Artificial Intelligence, Neural Networks, Multimedia and Computer Software