Load Balancing Optimization in Cloud Computing: Applying Endocrine-Particale Swarm Optimization

Shahrzad Aslanzadeh
Department of Engineering and IT
Center of Real Time Information Network
University of Technology Sydney
Sydney, Australia
Shahrzad.aslanzadeh@student.uts.edu.au

Zenon Chaczko
Department of Engineering and IT
Center of Real Time Information Network
University of Technology Sydney
Sydney, Australia
Zenon.chaczk@uts.edu.au

Load balancing optimization is categorized as NP-hard problem, playing an important role in enhancing the cloud utilization.

Different methods have been proposed for achieving the system load balancing in cloud environment. VM migration is one of these techniques, proposed to improve the VMs' functionality. Despite of the advantageous of VM migration, there are still some drawbacks which urged researchers to improve VM migration methods.

In this paper we propose a new load balancing technique, using Endocrine algorithm which is inspired from regulation behavior of human's hormone system. Our proposed algorithm achieves system load balancing by applying self-organizing method between overloaded VMs. This technique is structured based on communications between VMs. It helps the overloaded VMs to transfer their extra tasks to another under-loaded VM by applying the enhanced feed backing approach using Particle Swarm Optimization (PSO).

To evaluate our proposed algorithm, we expanded the cloud simulation tool (Cloudsim) which is developed by University of Melbourne. The simulation result proves that our proposed load balancing approach significantly decreases the timespan compared to traditional load balancing techniques. Moreover it increases the Quality Of Service (QOS) as it minimizes the VMs' downtime.

Keywords- Cloud computing, load balancing, task scheduling, Hormon regualtion, Endocrine Algorithm

I. INTRODUCTION

Significant characteristics of cloud computing such as elasticity, scalability and payment model attract businesses to replace their legacy infrastructure with the newly offered cloud technologies. As the number of the cloud users is growing rapidly, extensive load volume will affect performance and operation of the cloud. Therefore, it is essential to develop smarter load management methods to ensure effective task scheduling with efficient resource management. [1]

By applying the virtualization techniques, cloud computing is able to offer different services such as infrastructure as a service (IaaS), Platform as a service (PaaS) and Software as a service (SaaS).[2] Moreover virtualization provides servers with dynamic load balancing, monitoring the allocation and release of the resources. Despite of the advantageous of this method there exist some drawbacks

related to overloading status of the VMs. To solve this issue, VM-migration was suggested by researchers to re-map the overloaded VMs with available resources which have enough processing power to complete the tasks' executions. The main difficulty of applying this method is that the overloaded VMs will be migrated to a new host, employing the large amount of system's memory from both overloaded and the under loaded host. Therefore huge memory will be wasted while VM's functionality freezes as it cannot offer any more services. [3]

Hence in this paper we propose a self-organizing optimization method which achieves system load balancing base on VMs' communications records. Applying this approach, overloaded VMs transfer their extra tasks to available neighbors using feed backing method. To find the optimal VM for allocating the extra tasks, we selected the PSO algorithm as an optimization technique. To evaluate the model we extended the Cloudsim package incorporated with improved Jswarm module.

The remaining of this paper is organized as follows. In section II related works in load balancing optimization is discussed. Section III presents the problem formulation followed by mathematical concept of Endocrine-PSO algorithm in part IV. Section V illustrates details of our proposed method. Section VI explains the evaluated results. Finally section VII presents the conclusion and the future work of our research.

II. RELATED WORKS IN LOAD BALANCING OPTIMIZATION

Considering VM migration, several methods have been proposed which freeze the overloaded VM by migrating the whole VM to available new host with required processing power. Employing these methods, the migrated services are halted until memory resumes on new powerful host.

Tsakalozos [4] proposed a new approach which is applying a special purpose file system for performing the VM replication. In these methods, brokers are designed to monitor the migration process and avoid the network congestion by up to 24%.

Weida [5] presented a method which considers the VM's downtime and VM's speed during the migration process.

Kang [6] also recommended a new feedback based method which considers the total migration time. The goal of their

method is to improve the sequential VM migration time without affecting the performance of the multiple hosts.

Wolke[7] in their work designed a new software called DSAP which calculates the VM allocations and live migrations to architect different load patterns. They showed that dynamic allocation can improve the server performance while enhancing the quality of service.

Deshpande [8]offered a new method called scatter-Gather which is minimizing the eviction time by applying the pre/post -copy method. In this method they tryied to decouple the source and the destination by scattering the VM's memory state to midlle boxes. Then the destination source would gather the VMs' memory state by accessing the middle boxs which is completing the post-copy action.

Lin et al [9] presented a new method which is monitoring the load changes between the VMs and then according to the results automatically determines the tasks that should be migrated from selected VMs.

Li et al [10] applied a resource allocation technique which is composed of dynamic min-min and list scheduling algorithm. This method is architected for IaaS layer and works based on the priority of the tasks' execution times by monitoring the resource status.

Migrating the complete VMs and urging VMs to shut down for un-known time frame is the main drawback of the reviewed methods. To improve the VM migration method, we are proposing a new load balancing technique which is inspired from regulation of the Human's hormone system, called Endocrine algorithm. This approach enables overloaded VMs to transfer their extra tasks to available VM via communications which is inspired from feed-backing concept between hormones. Considering the benefits of this method, there is no urged shutdown for VM's applications during the migration frame and memory will not be wasted for any unnecessary pre-copy process. As a result the QOS will be enhanced dramatically.

III. PROBLEM FORMULATION

To model our self-organizing task scheduling method, we improved the techniques proposed in [11,12] for formulating the task execution time. In this method, the goal is to minimize the spantime which is total execution time for submitted tasks based on the load projections of the VMs. To formulate the model, we determined

W_i $i = \{W_1, W_2, \ldots, W_n\}$ as n dependent workflow tasks; CP_j j={1,2, ….m} as m computing power; k defines the numbers of the VMs where VM_t $t = \{1,2 \ldots k\}$ and B_s define the bandwidth between two VM_k. It should be noted that $X_{ik} = 1$ if task i is assigned to VM_t otherwise $X_{ik} = 0$. DE_{ik} defines the amount of data that task i assigns to VM_t. VM_m and VM_c represents the memory and capacity of the VM_k. Based on the defined parameters, we can denote the execution time on VM_k as:

$$Texe_k = \sum_{i=1}^{n} \sum_{k=1}^{m} X_{ik} * \frac{DE_{ik}}{VM_m * VM_c} \quad (1)$$

Where the total task execution time is:

$$T_{exe} = \sum_{k=1}^{m} Texe_k \quad (2)$$

Then the problem could be defined as:

$$\min f(Time) = (T_{exe} + (max_{i=1}^{k} Texe_k)) \ (3)$$

Subject to:

$$\sum_{k=1}^{m} X_{ik} = 1, \forall\ i = 1, \ldots, n \quad (4)$$

$$X_{ik} \in \{0, 1\}, \forall\ i = 1, \ldots, n\ , k = 1, \ldots, m$$

IV. MATHEMATICAL MODEL OF OUR PROPOSED ALGORITHM

The main elements of endocrine systems are hormones and glands. The information in human body is released by glands and transferred by hormones from one cell to another. Cells will not accept all the hormones. Each hormone is genetically programmed in a way that only can be received by certain cells.

The most important glands in human body are hypothalamus and pituitary which produce the vital hormones of the human body. Moreover they control different glands in endocrine system. By using the feed-backing function, these two glands will balance the hormones' load in human body.

Once the gland releases the hormone, it will travel in bloodstream until it reaches its target cell. Each cell has its own receptors. Receptors are responsible for communicating with hormones to ensure only specific hormones are received by the target cell.

To describe the mathematical model of the endocrine system three different models are considered [13]:

a) Single gland: In this model one gland is intracting with itself to reach the stable state. The mathematical model for this architecture can be described in equation (4) G_1, G_2, G_3 refer to the single gland:
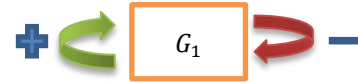

Figure 1- Hormone regulation with single gland

$$G_1 = ((a - b) * G_1) * G_1 \quad (4)$$

b) The second model is known as Lotka-Volterra which depicts two different intracting glands. The two glands are intracting with themslevs and each other.
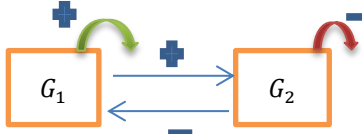

Figure 2- Hormone regulation with two glands

$$G_1 = (a - (b * G_2)) * G_1 \quad a,b,c,d > 0 \quad (5)$$

$$G_2 = (c - (d * G_1)) * G_2 \quad (6)$$

c) In third model, three glands communicate with themselves to control the regulation of the released hormons.
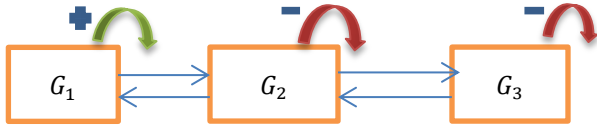

Figure 3- Hormone regulation with three glands

$$G_1 = (a - (b * G_2) - (c * G_3)) * G_1 \quad (7)$$

$$G_2 = (-d + (e * G_1) - (f * G_3)) * G_2 \quad (8)$$

$$G_3 = (-g + (h * G_2) - (i * G_1)) * G_3 \quad (9)$$

A,b,c,d,e,f,g,h,i>0

Inspiring from the above described models, in our solution, Glands are representing the VMs. We apply the principal of hormones' regulation to migrate the tasks from overloaded VM to another available VM. Therefore, to find the best available VM to complete the migration we have applied PSO to find the best trade-off between available VMs.

In robotics science, Endocrine system plays a key role. The feed backing methods between hormones have been applied many times to control the behavior and reactions of the robots. Also, researchers have improved the task assignment performance in robots by modeling the Endocrine regulation system. [14]

In this paper, we are combining the endocrine regulation with PSO functions to find the best possible trade-off among the Virtual machines.

PSO is a population based heuristic algorithm. By applying PSO for VM monitoring, each VM will be a potential solution in task scheduling.

Each VM has a position vector (P_k) and a velocity vector (L_i). In each migration, the kth VM looks for the best position that it could obtain to complete the migration (VM_{bestk}) among other existing VMs (VM_{exist}).

Therefore, to update the velocity and position of the VMs following equations should be considered.

Velocity equation:

$$L_i(t + 1) = WL_i(t) + C_1r_1(VM_{bestk}\text{-}P_k(t)) + C_2r_2(VM_{bestk}(t) - (P_k(t)) \quad (10)$$

Position equation:

$$VM_i(t + 1) = VM_i(t) + VM_i(t + 1) \quad (11)$$

In these equations C_1 and C_2 are positive constant. And r_1 and r_2 are random values between [0,1] and W is the inertia weight. For each VM, VM_{bestk} values are compared and updated. The theory behind the Endocrine-PSO algorithm is based on push-pull behavior between glands. The term "push-pull" indicates stimulating (push) a hormone from one gland and inhibition (pulling) hormones to other gland in closed loop. Cleary "push-pulling' function is highlighting the phenomenon of hormones regulation.

V.   ALGORITHM DESCRIPTION

We established our Endocrine-PSO based algorithm with objective of minimizing the timespan. We assume that we have n tasks that should be assigned to k VMs.

In Endocrine-PSO algorithm we need to consider two main particles for performing the pushing and pulling functions. One for simulating hormones (S_t) and the other for inhibiting the hormones (D_t).The position and velocity of the particles are set randomly. Stimulating hormones initialize (S_t) and inhibiting hormones initialize (D_t).

Next, according to [15], S_t and D_t are grouped to create class of U_t. Then, the non-dominiated solutions of U_t will be selected as a new S_t in t+1 time. In next step, the crowding distance of S_{t+1} is calculated and ranked in descending order.[16]

Finally, in last step, best position of the particle is determined (X_{Pbesti}) which respectivitly derives the global position of the particle (X_{gbesti}). The Pseudo-code of the algorithm is described in table 1.

VI.   SIMULATION RESULTS

We implemented our solution using cloudsim simulator. We improved Jswarm package to Endocrine-Jswarm to combine the PSO algorithm with Endocrine model. To evaluate our proposed method, the experiment completed with 5 VMs and 10 tasks which are submitted to overloaded VMs. Tasks' sizes are equal and each of them need 1 CPU to be completed. The details of the VMs and tasks are captured in Table.2 and Figure.4. Using cloudsim, our proposed algorithm managed to move the extra tasks from overloaded VMs to more powerful VM, aiming to complete the process with minimized timespan.

According to the output results, tasks (0,1,2,3,4,5,6,7,8,9) were allocated to VMs(4,5,3,3,1,1,3,5,3,4) which is shown in Figure.4.As Walker [17] noted, the minimum time needed

Table 1- Pseudo code of Endocrine-PSO based algorithm.

Step1. Initialize position and velocity of the particles in swarm population. Initialize Stimulating hormones S_t and inhibiting hormones D_t randomly.
Step2. Initialize the non-dominated solutions by merging D_t and S_t to class of U_t.
Step 3. Calculate the best positions of the (X_{pbesti}) in class of U_t.
$C(S_i(t))=\{C\epsilon D_t$, dist $(c,S_i(t))=\min(dis(c, S_j(t)))\}$,j=1.2.sizeS
$X_{cbesti}(t) = S_i(t)$ if individual i∈ $C(S_i)$
Step 4. Calculate the DE_{ik}, VMm_k, VMc_k, and Bc_k based on the united S_t and D_t group.
Step 5. Evaluate population according to defined fitness (min exe time) using equation (2) and (3)
Step 6. Sort the value of crowding distance in descending order
Step 7. Calculate (X_{pbesti})and (X_{gbesti}) update velocity and position
Step9. Set t=t+1
Step 10. If termination satisfied, X_{gbesti} is the best task scheduling path
    ELSE
       Go to step 3

-to achieve load balancing with traditional algorithms is about 5 minutes. The results of our proposed algorithm showed that after 100 iterations, load balancing could be achieved in 50.114s which is less time consuming than traditional load balancing algorithms. Moreover, in traditional load balancing approaches, the entire overloaded VM should be migrated to the new VM, while with our proposed approach based on communication and feed backing between VMs only extra tasks should be migrated to new VMs which significantly reduces the memory consumption.

Table 2-VMs characteristics

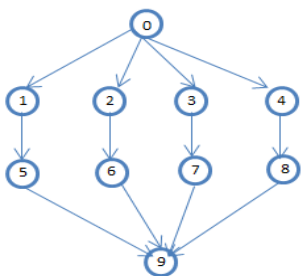| VM Id	MIPS	Bandwidth	CPU
0	256	10000	4
1	256	10000	2
2	256	1000	4
3	256	1000	1
4	256	10000	3

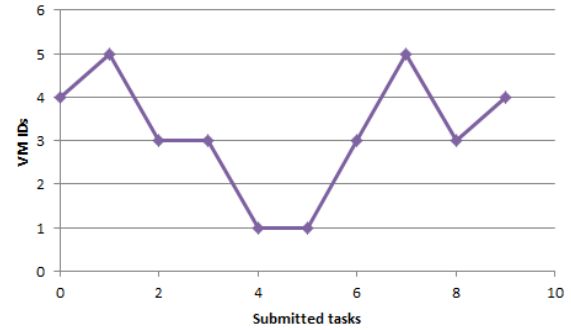Figure 4- Structure of the submitted tasks

Figure 5- Selected VMs for completing the VM migration

VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new task scheduling algorithm which is inspired from regulation of the hormones in endocrine system. We designed an algorithm to minimize the timespan during the VM migration.
To evaluate our proposed method we extended Cloudsim simulation tool, which is an open source cloud simulator. The experimental results in simulation environment show that Endocrine-PSO offers the best position for migrating the tasks from overloaded VM to more powerful VM. Regulation of the endocrine system was the main idea of this algorithm which plays an important role in evolutionary computing methods. This algorithm can help data centers to achieve the desired load balancing in in their systems.

References

[1] Rimal, B., E. Choi, et al. (2010). A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems. Cloud Computing. N. Antonopoulos and L. Gillam, Springer London: 21-46.

[2] Gupta, A. Cloud computing growing interest and related concerns. Computer Technology and Development (ICCTD), 2010 2nd International Conference on, 2-4 Nov. 2010 2010. pp. 462-465

[3] Prakash Kumar, Pradeep Kumar and Vikas Kumar. Article: Load Balancing Strategy for Computational Grid System using an Improved Algorithmic Approach. International Journal of Computer Applications 79(8):46-50, October 2013. Published by Foundation of Computer Science, New York, USA

[4] Tsakalozos, K., et al. (2014). Time-Constrained Live VM Migration in Share-Nothing IaaS-Clouds. Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on.

[5] Weida, Z., et al. (2014). Adaptive Live VM Migration over a WAN: Modeling and Implementation. Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on.

[6] Kang, T. S., et al. (2014). Design and Implementation of Middleware for Cloud Disaster Recovery via Virtual Machine Migration Management. Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on.

[7] Wolke, A. and L. Ziegler (2014). Evaluating Dynamic Resource Allocation Strategies in Virtualized Data Centers. Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on.

[8] Deshpande, U., et al. (2014). Fast Server Deprovisioning through Scatter-Gather Live Migration of Virtual Machines. Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on

[9] Lin, W., Wang, J.Z., Liang, C., Qi, D.: A threshold-based dynamic resource allocation scheme for cloud computing. Procedia Eng. 23, 695-703 (2011)

[10] Li, J., et al. (2012). "Online optimization for scheduling preemptable tasks on IaaS cloud systems." Journal of Parallel and Distributed Computing **72**(5): 666-677.

[11] Guo,L., Zhao, S., Shen, S., Jiang, C.: Task scheduling optimization in cloud computing based on heuristic algorithm. Journal of Networks 7(3), 547-553 (201`2)

[12] Liu, H./ Abraham, A., Snasel, V., McLoone, S.: Swarm scheduling approaches for workflow applications with security constraints in distributed data-intensive computing environment. Information Science 192(0), 228-23 (2012)

[13] Koos,I, Mathematical models of endocrine systems. Health faculty collage of Hungary,2011

[14] M.Mendao,A neuro-endocrine control architecture applied to mobile robotics.PhD thesis, university of Kent, Canterbury, UK, 2007

[15] K.Testsuo, N.Eiji, M.Peter, Hormone evolution: the key to signaling, Nature (422) (2003)122.

[16] C.R.Raquel, P.C. Naval, An effective use of crowding distance in multiobjective particle swarm optimization, in proceeding of genetic and Evolutionary computation Conference (GECCO 2005), Washington, D.C., June 25-29, 2005

[17] J.Walker, M. Wilson, A performance sensitive hormone-inspired system for task distribution amongst evolving robots, in: intelligent Robots and systems, 2008. IROS 2008. IEEE/RSJ International Conference on 2008, 9, 22-26, pp.1293-1298, 2008.