Journal of Enterprise Information Management

Service value optimization of cloud hosted systems using particle swarm technique

Salah Eddin Murad Salah Dowaji

Received 18 April 2015
Revised 14 August 2015
23 October 2015
26 October 2015
Accepted 2 November 2015

Service value optimization of cloud hosted systems using particle swarm technique

Salah Eddin Murad and Salah Dowaji
Department of Software Engineering and Information Systems, Damascus University, Damascus, Syria

Abstract

Purpose – Cloud Computing has become a more promising technology with potential opportunities, through reducing the high cost of running the traditional business applications and by leading to new business models. Nonetheless, this technology is fraught with many challenges. From a Software as a Service (SaaS) provider perspective, deployment choices are one of the major perplexing issues in determining the degree to which the application owners' objectives are met while considering their customers' targets. The purpose of this paper is to present a new model that allows the service owner to optimize the resources selection based on defined metrics when responding to many customers' with various priorities.

Design/methodology/approach – More than 65 academic papers have been collected, a short list of the most related 35 papers have been reviewed, in addition to assessing the functionality of major cloud systems. A potential set of techniques has been investigated to determine the most appropriate ones. Moreover, a new model has been built and a study of different simulation platforms has been conducted.

Findings – The findings demonstrate that serving many SaaS customer requests, with different agreements and expected outcomes, would have mutual influence that impact the overall provider objectives. Furthermore, this paper investigates how tagging those customers with various priorities, with reflection of their importance to the provider, permits controlling and aligning the selection of computing resources as per the current objectives and defined priorities.

Research limitations/implications – This study provides researchers with a useful literature, which can assist them in relevant subject. Additionally, it uses a value-based approach and particle swarm technique to model and solve the optimization of the computing resource selection, considering different business objectives for both stakeholders, providers and customers. This study derives priority of a number of factors, by which service providers can make strong and adaptive decisions.

Practical implications – The paper includes implications on how the SaaS service provider can make decisions to select the needed virtual machines type driven by his own preferences.

Originality/value – This paper rests on the usage of Particle Swarm Optimization technique to optimize the business value of the service provider, as well as the usage of value-based approach. This will help model that value in order to combine the total profit of the provider and the customer satisfaction, based on the agreed budget and processing time requested by the customer. Another additional approach has been charted by using the customer severity factor that allows the provider to reflect the customer importance while making the placement decision.

Keywords Cloud Computing, Particle Swarm Optimization, Business driven IS management, SaaS management, Value as a service

Paper type Research paper

1. Introduction

From a business perspective, Cloud Computing, which presents a new paradigm for distributed computing, is about improving organizational efficiency and reducing operational cost, often coupled with the objective of achieving a faster time to market (Hauck *et al.*, 2010). Cloud Computing delivers three defined models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

Journal of Enterprise Information Management
Vol. 29 No. 4, 2016
pp. 612-626

1741-0398
DOI 10.1108/JEIM-02-2015-0018

SaaS refers to prebuilt, functionally independent, vertically integrated and universally available applications delivered to and used by customers as services. PaaS refers to software and product development, which clients lease so they can build and deploy their own applications for their specific use. IaaS is essentially hardware devices and generally refers to a virtualization environment where services enable the Cloud platforms and applications to connect and operate.

In Cloud systems, it is critical for cloud providers to take influential decisions to make the most suitable deployment choices in order to maintain the required and acceptable value of his/her objectives. Achieving the required value needs a high level of adaptability and responsiveness to the business requirements and objectives. This mission becomes more complicated when serving different customers with different agreements, hence with various expected outcomes.

This paper presents a new model that allows the service owner to optimize the resources selection based on defined metrics when responding to many customers' with various priorities. The paper rests on the usage of Particle Swarm Optimization (PSO) technique to optimize the business value of the service provider, as well as the usage of value-based approach. PSO, an applied soft computing method developed by Kennedy and Eberhart (1995), is one of the most advance evolutionary algorithms driven from nature. PSO approximates an optimal solution by iteratively improving a group of candidate solutions, called particles. PSO has become popular due to its simplicity and its effectiveness in a wide range of applications with low computational cost. The business objectives, on the other hand, are modelled using value-based approach presented in our previous work for SaaS management, called VBM (Murad and Dowaji, 2014). It allows modelling those objectives in alignment with varying required metrics of different weighted customers.

The remainder of the paper is structured as follows: Section 2 presents the related works. Section 3 outlines the research methodology. Section 4 defines the problem of business value maximization. The evaluation and experiments are presented in Section 5. Section 6 is dedicated to research implications and limitations while Section 7 is the summary and futuristic works.

2. Related works

Different available service models of Cloud Computing paradigm have made it more feasible for service providers to deliver their services as a utility to the customers in different methods. This is mainly and specifically due to: virtualization and SaaS model.

2.1 Virtualization

Virtualization is a process of substitution in which a physical resource is replaced by many logical (virtual) resources (Litoiu *et al.*, 2010). Organizations, therefore, can govern and manage cloud services as a critical element of running IT as a business.

2.2 SaaS

SaaS is the delivery of software functionality online, which is similar to the one installed on a local machine. Depending on the content of the service, a cloud can offer IaaS, PaaS and SaaS (Litoiu *et al.*, 2010). The services that are provided by Google for office automation is one of SaaS examples, this includes Google Mail, Google Documents and Google Calendar. SalesForce.com is a provider of commercial solution which provides online customer relationship management. Another commercial solution provider is Clarizen.com which provides a project management services.

Appirio is an integrated solution that provides complete support for any management aspect of modern enterprises from project management to resource planning. Appirio is able to be integrated to the platform additional services exposed by other Clouds such as Amazon EC2, SalesForce.com, Google AppEngine and Facebook (Buyya *et al.*, 2013).

This computational model, Cloud Computing, has been presented in many works, which have been assumed to deal with different issues. There have been a number of attempts that focused on the technical perspective of the computing model including performance enhancement (Alsolami, 2013; Roy and Dutta, 2013), availability (Birman *et al.*, 2009), data storage (Gheorghe *et al.*, 2010; Pretschner *et al.*, 2008) and privacy (Unruh and Müller-Quade, 2010; Künzler *et al.*, 2009; Dowsley *et al.*, 2009). Other researches ranged from computer networking, distributed file systems, distributed database to computational job scheduling problems (Chun and Culler, 2002; Coleman *et al.*, 2004).

On the other hand, different aspects of business influencing factors have been studied in many other endeavours due to the fact that businesses now expect that IT cloud services and infrastructures should bring them closer to achieve their own business objectives. Chauhan (2011) has provided a general framework for vendor-independent cloud management, which chooses and executes cloud management actions that are optimal businesswise. He extends the WS-Policy4MASC to support the management actions and metrics.

Elmroth *et al.* (2009) proposes an accounting and billing architecture between the consumers and the infrastructure provider in a federated cloud infrastructure. The objective is to cope with the migration of virtual machines (VMs) by managing pre-paid and post-paid payment schemes according to the users' needs.

An architecture for business driven IT management model has been proposed by Oriol and Guitart (2011). It presents three layers: cloud layer, business layer and the execution layer. Adopting this model by the cloud providers allows for business level governance. Litoiu *et al.* (2010) have discussed several facets of optimization in Cloud Computing, proposing architecture for corresponding challenges. It considers a layered cloud to support the self-managed cloud service.

Moreover, many endeavours have addressed issues of resource allocation in cloud environment. Sauvé *et al.* (2005) have proposed a model that realizes the best service value for the provider. However, their model only considers the business losses due to IT service failure or performance degradation, using a static method for resource allocation.

Li and Chinneck (2009) have presented a method for achieving optimization in clouds, using performance models in the development, deployment and operation of the cloud-hosted application. Nonetheless, this optimization model only considers the cost reduction target, using the response time to determine the Quality of Service (QoS) presented to the customer.

Li *et al.* (2013) introduce an approach for improving the availability guarantee of software applications by optimizing the availability considering the performance and cost. Their approach considers how sharing the VM resource with many users can affect the availability.

From a provider perspective, the minimization of the cost while meeting the Service-Level Agreement (SLA) has been deliberated by Karve *et al.* (2006). A management algorithm is proposed to maximize the utilization of active VMs and minimize the SLA violation. Furthermore, Yu and Buyya (2006) present a budget constraint scheduling of a workflow applications. They proposed a scheduling approach to minimize the execution time while meeting the specified budget for delivering the result. Khajemohammadi *et al.* (2014) used levelled multi-objectives Genetic Algorithm (GA) in order to reduce the scheduling time when running workflows on grid environment. The aim is to allocate each task in the workflow on the most appropriate resource in the grid based on the user requirements, cost and deadline.

Profit maximization of using cloud resources has been presented by Tsakalozos *et al.* (2011). Their approach answers the question of how many resources a consumer should request from the seemingly endless pool provided by the cloud. Although they model the budget constraint as a function of response time, they did not consider the different types of VM resources or price changes over time due to the competing for resources, without accounting for the waiting time.

The financial model presented by Dash *et al.* (2009) uses budget constraint to manage the cloud environment. Though the approach responds to user QoS requirements by building new data structures, it fails to adapt with the number of VMs deployed.

A new approach to schedule Bag of Tasks (BoTs) with increase in parallel process presented by Pingale and Mogal (2015). They proposed an algorithm for multi-objectives scheduling. The optimized objectives are the execution time, the network bandwidth and the storage requirements.

Alkhanak *et al.* (2015) discussed the challenges of workflow scheduling that affect the workflow cost. They classified the cost-aware relevant challenges of workflow scheduling in Cloud Computing depending on the QoS performance, system functionality and system architecture. Ultimately they provided a taxonomy set in order to facilitate the selection of the appropriate approach from the available alternatives.

According to Pandey *et al.* (2010), GAs have been used for scheduling problems in many other works. Ahmad *et al.* (2002) have shown the effectiveness of the PSO-based algorithm in comparison with GA on randomly generated task interaction graphs. Furthermore, the results of simulated experiments, presented by Zhang *et al.* (2008), show that the PSO algorithm is able to get the better schedule than GA in grid computing environments. In addition, the results presented by Liang *et al.* (2007) have provide evidence that PSO algorithm is able to improve 57 out of 90 best-known solutions provided by other well-known algorithms in solving the sequencing problems. For instance, PSO-based heuristic approach for scheduling workflow applications appears in work presented by Pandey *et al.* (2010). This approach takes into account both the computation cost and data transmission cost of applications, while seeking the minimization of the total cost.

Our participation in this paper hinges on the usage of PSO technique to optimize the business value of the service provider. This value is modelled by using a value-based approach, presented in our previous work (Murad and Dowaji, 2014). The study aims to enable the application owners to optimize the value of the service they provide by managing and controlling the selection the most appropriate computing resources directed by some metrics. Therefore, this paper combines the total profit of the provider and the customer satisfaction based on agreed budget and processing time requested by the customer, using the customer severity factor that allows the provider to reflect the customer importance when making the placement decision. The priorities of the customers are derived on the basis of the SLA that governs the relation between the two parties and can be driven, with the other priorities of total profit and customer satisfaction, by a set of key performance indexes (KPIs) that specify his/her priorities at the current stage. The PSO-based algorithm, used to make resource allocation decisions, considers different kinds of resources and changed prices of the potential cloud resources as well.

2.3 PSO

PSO is one of the evolutionary algorithms inspired by the social behaviour of fish schooling or bird flocking (Ahmad *et al.*, 2002). A particle in PSO is analogous to a bird or fish flying through a search (problem) space. In the PSO algorithm, each particle represents a possible solution. The swarm of particles is initially randomly generated. Each particle has its own position in the space and a fitness value, and has the velocity to determine the speed and direction it flies. A group of candidate solutions (particles) moves around in the search space based on the particles' updated position and velocity so that the PSO algorithm can get an optimized solution.

Particles in the search process update themselves by tracking two best-known positions. One best-known position known as local best position is the individual best-known position in terms of fitness value reached so far by the particle itself. Another best-known position, known as global best position, is the best position in the entire population (Wu, 2014). Suppose the number of particles is N. The velocity and position of each particle are calculated by formulations (1) and (2):

$$v_i^{k+1} = wv_i^k + c_1 r_1 \left(p_i^k - x_i^k\right) + c_2 r_2 \left(p_g^k - x_i^k\right) \tag{1}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{2}$$

$$p_i^k = \text{best}\left(x_i^k, p_i^{k-1}\right) \tag{3}$$

where v_i^k is the velocity of particle i at iteration k; v_i^{k+1} the velocity of particle i at iteration $k+1$; x_i^k the position of particle i at iteration k; x_i^{k+1} the position of particle i at iteration $k+1$; w the inertia weight; c_i the acceleration coefficients; $i=1, 2$; r_i the random number in [0,1]; $i=1, 2$; p_i^k the best position of particle i; and p_g^k the best position of particle i in all particles.

3. Research methodology

A procedure has been created to manage the research progress starting from an extensive search of the high-related keywords over the most important search engines and repositories like IEEE Xplore, Science Direct, Google Books and Google Scholar. A set of more than 65 academic papers have been collected, a short list of the most related 35 papers have been reviewed. It mainly included sources from books, journals, conference proceedings and available working papers related to Cloud Computing, Management of Cloud Computing, Business Driven Management and other related topics. The research is also based on the review of the functionality of the major cloud systems including Amazon EC2, Salesforce and Microsoft Azure.

A set of potential techniques including value metrics measurement, GAs, Pareto-Optimal algorithms and PSO has been identified and investigated to determine the most appropriate approach. A new model has been built and a study of the most adaptable simulation frameworks and tools has been carried out including, but not limited to following.

iCanCloud: a simulation platform aimed to model and simulates Cloud Computing systems, which is targeted to those users who deal closely with those kinds of systems. The main objective of iCanCloud is to predict the trade-offs between cost and performance of a given set of applications executed in a specific hardware, and then provide to users beneficial information about such costs (Castane *et al.*, 2012).

CloudSim: a Framework for modelling and simulation of Cloud Computing infrastructures and services (Calheiros *et al.*, 2011).

In this paper, CloudSim simulation framework is used to test and assess the proposed model. Repeated runs have been performed using different and randomly generated set of parameters that simulates different scenarios and cases. The average value of the outcomes has been used to evaluate the results.

The detailed steps followed throughout the research are illustrated by Figure 1.

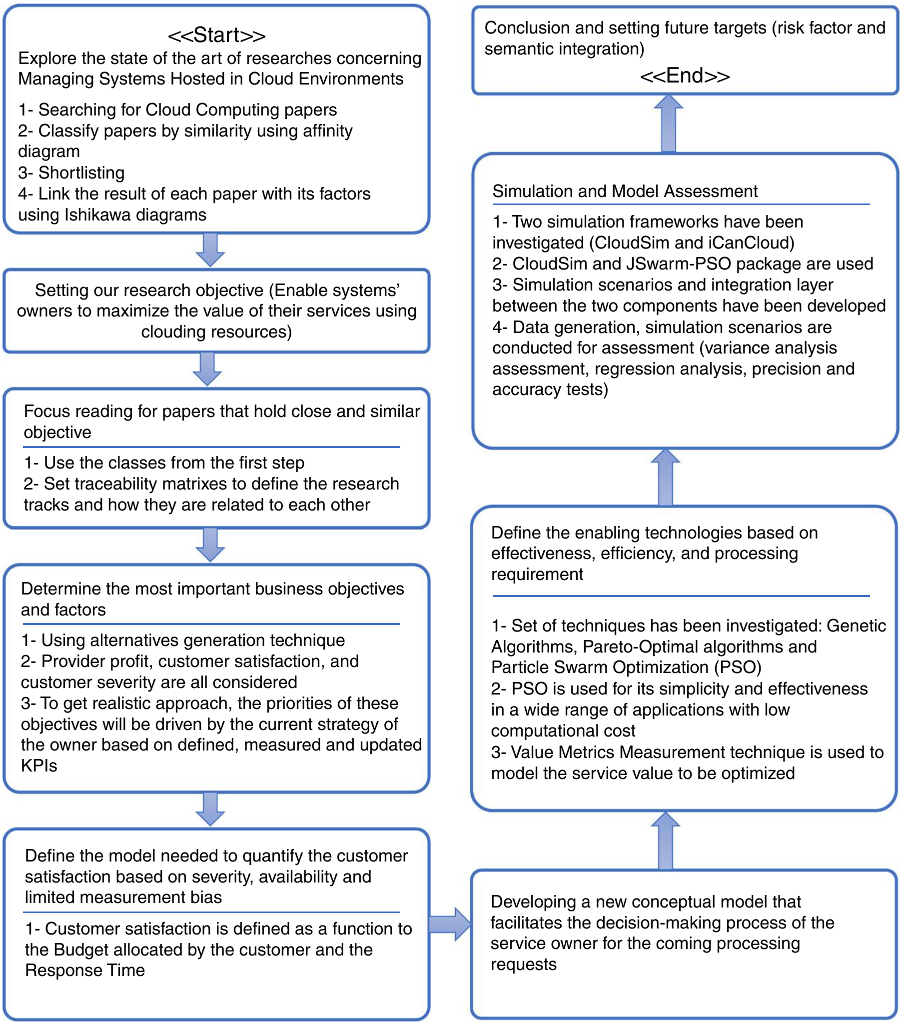

Figure 1.
Research procedure

4. Business value maximization

The application owner serving many customers may have multiple objectives to achieve. Total profit and customer satisfaction are example of these objectives. Focusing on one of those objectives may affect the value of one another, i.e. giving more priority to profit will decrease the customer request value. It could be undesired to influence those values for all of the customers to the same degree. Therefore, specifying the severity of the customer, according to the SLA signed between the two parties, is a critical factor that could be considered and defined by the provider when processing the clients' requests.

The business service provided here is an on demand application service in the cloud. The provider receives requests by the customer(s) and has to select and rent VM instances to be used for customer request execution. As illustrated in Figure 2, the timeframe in the scheduling process is divided into a number of time slots. The processing of the customer request is performed during those slots. The windows between those slots are the period in which the rent decision is taken by the system owner and the bidding for the selected VM types are executed. The processing request of the customer can be received at any time and it has to wait until the next time slot to be considered by the SaaS owner. This waiting time will be considered as a part of the request response time. The provider will charge the customer for all of the VM instances that were rented to serve his/her request(s) during that time slots. The provider model is inspired from the Amazon EC2 (2014) spot.

The problem for the service provider is to optimize the service value by selecting the suitable VM instances provided by infrastructure service provider to response to the customer requests.

The objective function of our approach, presented in formula (4), is to maximize the business service value of the provider by serving N requests, each is modelled using the tuple (*RS*, *BD*, *CP*, *PT*, *w*1, *w*2) (Murad and Dowaji, 2014):

$$Max\ ServiceValue = P_1 \sum_j Profit_j + P_2 \sum_j \delta_j \times CRV_j;$$

$$j = \{1, 2, 3 \ldots N\}, P_1 + P_2 = 100 \qquad (4)$$

where P_1 and P_2 represent the current priority of the provider between the total profit and the customer satisfaction. $Profit_j$ is the profit of serving the request j using a certain VM instance, CRV_j is the customer request value that presents the customer

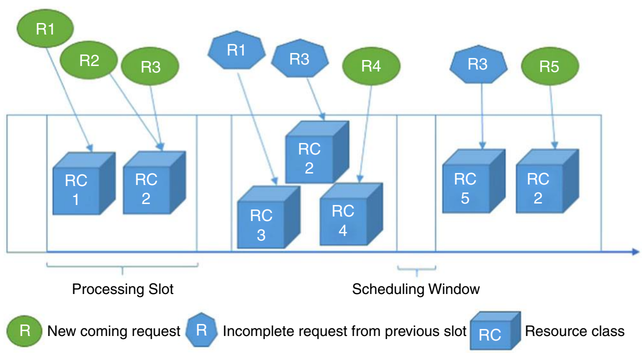

Figure 2.
Scheduling process timeframe

satisfaction of processing his\her request j on that instance. Formulas (5)-(7) present how *Profit* and *CRV* are calculated (Murad and Dowaji, 2014). δ_j is the severity factor of each customer to the provider where $\Sigma\delta = 100$:

$$CRV = W_1 \times BD + W_2 \times PT \tag{5}$$

$$Profit = Revenue–TotalCost \tag{6}$$

$$Revenue = \begin{cases} CP & \text{CRV} \geqslant 2.5 \\ Budget & 1.5 \leqslant \text{CRV} < 2.5 \\ Budget - pn & 1 \leqslant \text{CRV} < 1.5 \\ Budget - 2 \times pn & \text{CRV} < 1 \end{cases} \tag{7}$$

PSO particle is represented in Figure 3, which presents N dimensional particle, where each position represents a request and the values of each position is its mapping to a computing resource.

Driven by the PSO technique, the optimization process of selecting the suitable computing resources in our model uses two algorithms: the scheduling algorithm, Algorithm 1, and PSO algorithm, Algorithm 2, as listed next:

Algorithm 1-Scheduling

```
//R listofType Request, RC listofType ResourceClass
if R < > ∅
BestPosition = EvolvePSO(R)
Repeat
For each dim in BestPosition.dimensions
   SelectedRC = RentVMOfInstancesType (dim.InsanceType)
   Assign R.GetRequest(dim.Index) to SelectedRC for execution
End for
For all requests in process
   UpdateAccumulatedCosts()
   UpdateRemainingRequestsSize()
End for
   ReceiveNewComingRequests
   BestPosition = EvolvePSO(R)
Until no more requests to process
Else
   return Null
```

Algorithm 2-EvolvePSO

```
//R listofType Request, RC listofType ResourceClass
// 1. Initiation
Particle.dimension = R.Size
For each p in Particles
   p.positions = RandomSelect(RC)
   p.velocity = CreateRandomVelocity()
End for
// 2. Iterations
For each p in particles
```

New_pbest = p.CalculateFitnessValue() // using formula 4
If Newpbest > p.previous_best_pbest
  p.fitness_value = New_pbest
End for
Gbest = GetBestParticle(particles)
Foreach p in particles
  CalculateVelocity() // using Equation 1
  UpdatePosition() // using Equation 2
End for
If Not (stoppage criteria or maximum_iteration)
  Go back to the Iterations Label " Step 2".

An illustration of the high level interaction of the model's parts is presented in Figure 4.

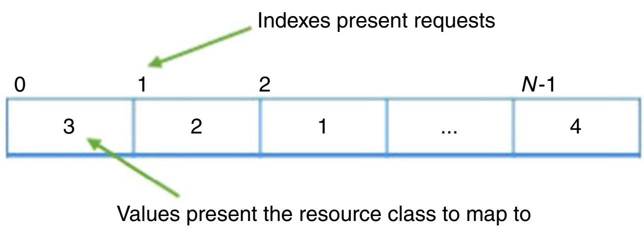

Figure 3.
PSO particle representation

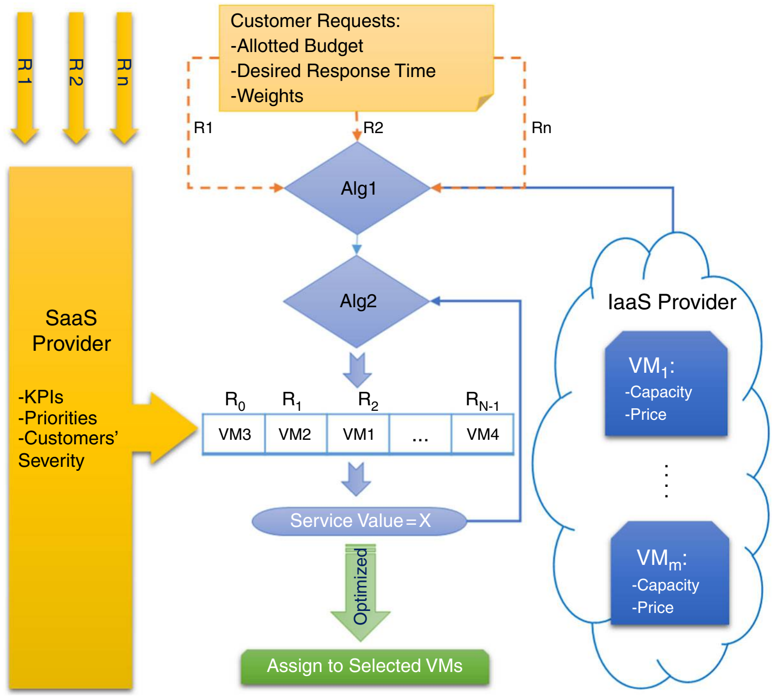

Figure 4.
High-level illustration of model interaction

5. Evaluation and experiments

An evaluation of our model and algorithms has been conducted based on Amazon EC2 (2014) instances, using CloudSim which is a framework for modelling and simulation of Cloud Computing infrastructures and services (Calheiros *et al.*, 2011).

The experiments aim to check the behaviour of our approach when using different parameters, priorities, computing capacities and prices of instances of different VM types.

5.1 The importance of customer severity factor

To present the effect of the customer severity factor to the provider, an assessment has been carried out using the parameters shown in Table I. Figure 5 shows how the value presented to the customer, customer request value, gets higher when using greater factor. This shows that this factor provide a control point for the service provider to mitigate the negative impact on the most important consumers.

5.2 The effectiveness of total profit priority and the customer importance on the customer request value

Predilection of the total profit of serving a set of requests over the customer satisfaction will drive to lower customer satisfaction, i.e. lower CRV. In this simulation, we aim to check how defining a different customer importance value will succeed to mitigate the negative impact on CRV for the most important customers. The result of this evaluation is plotted in Figure 6, where it clearly shows that the greater the severity of the customer the lower the differences in their CRVs. CRV difference metric measures the difference of the CRV for the same customer when the provider moves his/her priority from the customer satisfaction to the total profit.

Parameter	Value
No. of runs	5
Request size	600-1,500 MIPS
No. of requests	2,000 (25 per round)
No. of iterations	20
No. of particles	25
Instances	Small, large, extra, super
Instance prices	Amazon spot instances price history
Customer severity factor	Random [1,100]

Table I. Simulation parameters

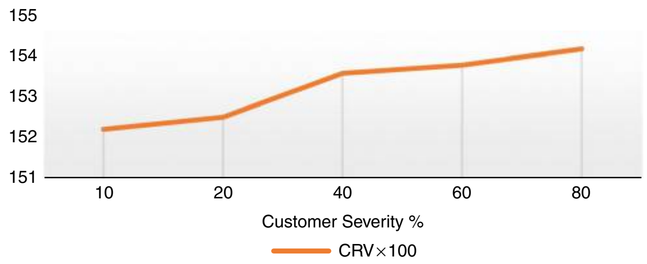

Figure 5. The importance of customer severity factor

5.3 Algorithm computing time in relation to the number of tasks

It is important to measure the time needed for the scheduling as it will show the feasibility of the presented approach. Therefore, this simulation is to examine the average time needed for the placement algorithm to run in relation to the number of requests to be served. For ten times repeated simulation we have spotted the processing time that is illustrated in Figure 7. This chart shows that it needs less than 350 ms to execute when having about 1,500 requests to optimally allocate resources for.

5.4 The relation between the total profit and profit priority

This section examines how different priorities of provider objectives affect the total profit and the value offered to the customer. For this purpose we have used randomly generated priorities to validate how those objectives go up and down driven by the provider preferences as proved in Figures 8 and 9. The first chart clearly shows how major growth in profit can be achieved when profit priority exceed 50 per cent. While the second graph shows the positive correlation between the value presented to the customer and the priority associated with it.

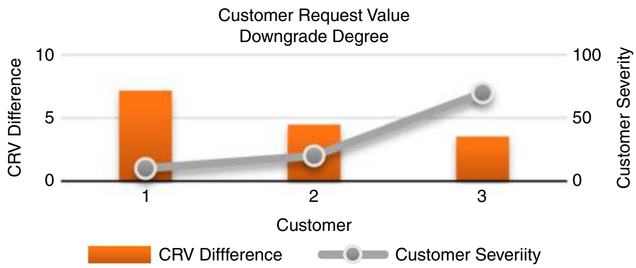

Figure 6. The effectiveness of total profit priority and the customer importance on CRV

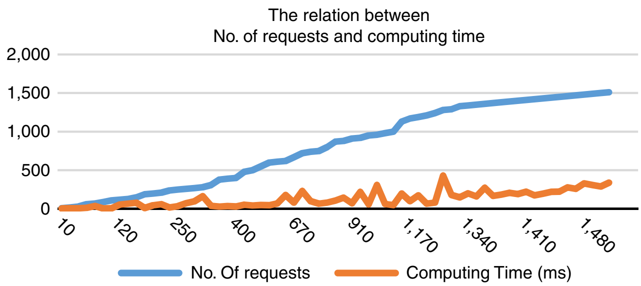

Figure 7. Algorithm computing time in relation to the number of tasks

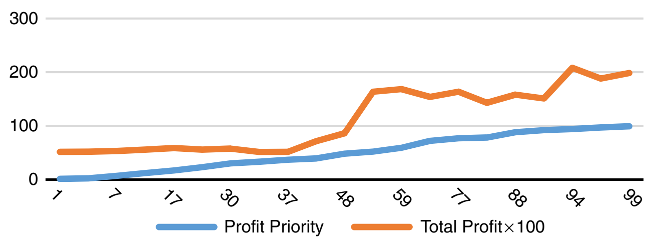

Figure 8. The relation between the total profit and profit priority

6. Research implications and limitations

Our simulation results provide several theoretical contributions in the area of management of information systems within Cloud Computing environments. For researchers, this study provides a useful literature, which can help them in relevant subject. It shows the effectiveness of the PSO-based model by optimizing the business value for the service owner.

The proposed framework can be used as a strategic tool that delivers competitive advantage to the organizations that deliver their services using Cloud Computing resources. It is practically a comprehensive framework since it covers the needed components to manage the whole process, from the start when receiving the processing request, right up until the provider delivers the service to the customer. This includes: how the service owner can determine the current strategy triggered by the readings of some live KPIs "which can follow an automated procedure", and how to reflect this strategy to the resource allocation decision making. Additionally, it enables the business owner to differentiate between profit and customer satisfaction, allows controlling the impact of the selected priorities based on the customer importance, allows measuring the value of the provided service and optimizes that value. Hereby, it provides a potential opportunity to successfully increase the benefits realization of the owned system. In this sense, the service owner is able to manage and control the computing resources, to be used, in more practical manner and in alignment with the desired strategy.

On the other hand, this research, along with a vast majority of the similar ones, assumes that the cloud infrastructure resources are available once needed. In this context, it does not consider the diversity of the prospective external resources including their characteristics and source "provider". Therefore, one of our developments will concentrate on investigating the addition of semantic web in order to deal with this issue.

7. Summary and future work

The adoption of Cloud Computing services results in new relations between the service providers and consumers as new business models and system architectures appear. Driven by their business needs and objectives, the providers require managing and controlling their applications in alignment with their most important KPIs. Consumers, on the other hand, have to specify their own needs per processing request. As an application owner, I need to set a decision-making structure that permits to manage my service driven by set of metrics considering those different requesters. In this work, we presented a PSO technique coupled with a value-based approach to optimize the

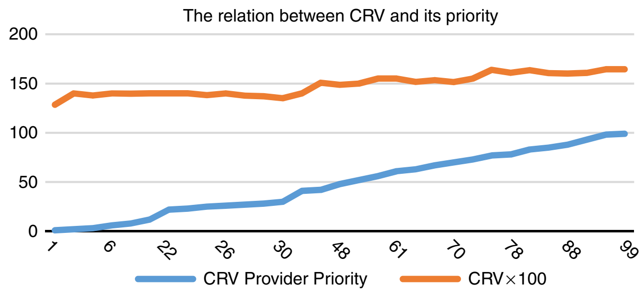

Figure 9.
The relation between the customer request value and customer satisfaction priority

selection process of infrastructure service provider resources, considering a set of defined business metrics of the application owner.

Our experiments demonstrate the relation between those objectives, provider and customer ones, and how the response to each customer request may affect the responses to the other requests and to the overall service value. Evaluation of using the customer severity factor has been performed and shows how it can help to differentiate between customers in relation to their importance to the provider. CloudSim tool has been used in those experiments using different and randomly generated parameters.

From practical perspective, this research is a further endeavour towards the motivation of the adoption of cloud service by providing the service owners with the technique to control and manage their own service in dynamic and flexible manner. This will encourage even the small and mid-size enterprises and providers to build and provide new services for public and to reach new customers with the least upfront investment and adapt their decisions in alignment to their strategy focus, i.e. profit, customer satisfaction and impact control.

Future work will focus on testing our model using real cases depending on actual computing resources and providers rather than simulation. Another research recommendation is to study how to find the potential set of infrastructure service providers based on some technical characteristics and many other attributes they can provide to the application owner regardless of having a predefined list of the computing resources providers and without human interference, semantic web can form an important addition in this regards. A selection model may be developed and tested to select the best provider, while decision-making criteria could be defined by the application owner to accept or reject the customer processing calls. Another factor to address is to account for the risks that are associated with using certain computing resources in comparison with other ones.

References

Ahmad, I., Salman, A. and Al-Madani, S. (2002), "Particle swarm optimization for task assignment problem", *Microprocessors and Microsystems*, Vol. 26 No. 8, pp. 363-371.

Alkhanak, E.N., Lee, S.P. and Khan, S.U.R. (2015), "Cost-aware challenges for workflow scheduling approaches in cloud computing environments", *Future Generation Computer Systems Journal*, Vol. 50 No. C, pp. 3-12.

Alsolami, F. (2013), "N-Cloud: improving performance and security in cloud storage", *IEEE 14th International Conference on High Performance Switching and Routing*, pp. 221-222.

Amazon EC2 (2014), *Instance Types*, available at: http://aws.amazon.com/ec2/instance-types (accessed 10 September 2014).

Birman, K., Chockler, G. and van Renesse, R. (2009), "Toward a cloud computing research agenda", *SIGACT News*, Vol. 40 No. 2, pp. 68-80.

Buyya, R., Pandey, S. and Vecchiola, C. (2013), "Market-oriented cloud computing and the cloudbus toolkit", in Sarbazi-Azad, H. and Zomaya, A.Y. (Eds), *Large Scale Network-Centric Distributed Systems*, ISBN: 978-0-4709-3688-7, Wiley Press, Hoboken, NJ, pp. 319-358.

Calheiros, R., Ranjan, R., Beloglazov, A., De Rose, C.A.F. and Buyya, R. (2011), "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", *Software: Practice and Experience (SPE)*, Vol. 41 No. 1, pp. 23-50.

Castane, G.G., Nunez, A. and Carretero, J. (2012), "iCanCloud: a brief architecture overview", *IEEE 10th International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, pp. 853-854.

Chauhan, D. (2011), "Business driven IT management in cloud computing systems", working paper, Sydney University, Sydney.

Chun, B.N. and Culler, D.E. (2002), "User-centric performance analysis of market-based cluster batch schedulers", *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID), May*, pp. 30-38.

Coleman, K., Norris, J., Candea, G. and Fox, A. (2004), "Oncall: defeating spikes with a free-market application cluster", *Proceedings of the 1st International Conference on Autonomic Computing*, pp. 198-205.

Dash, D., Kantere, V. and Ailamaki, A. (2009), "An economic model for self-tuned cloud caching", *Proceedings of the 25th IEEE International Conference on Data Engineering, Shanghai, March*, pp. 1687-1693.

Dowsley, R., Muller-Quade, J. and Nascimento, A.C.A. (2009), "A CCA2 secure public key encryption scheme based on the McEliece assumptions in the standard model", *CT-RSA*, pp. 240-251.

Elmroth, E., Marquez, F., Henriksson, D. and Ferrera, D. (2009), "Accounting and billing for federated cloud infrastructures", *Eighth International Conference on Grid and Cooperative Computing, GCC*, pp. 268-275.

Gheorghe, G., Neuhaus, S. and Crispo, B. (2010), "ESB: an enterprise service bus for access and usage control policy enforcement", *Proceedings Annual IFIP WG 11.11 International Conference on Trust Management*, Trento and Morioka, 16-18 June, pp. 63-78.

Hauck, M., Huber, M., Klems, M., Kounev, S., Pretschner, A., Reussner, R. and Tai, R. (2010), "Challenges and opportunities of cloud computing", technical report, Karlsruhe Institute of Technology, Wuerttember, pp. 1-11.

Karve, A., Kimbrel, T., Pacifici, G., Spreitzer, M., Steinder, M., Sviridenko, M. and Tantawi, A. (2006), "Dynamic placement for clustered web applications", *15th International Conference on World Wide Web, ACM*, pp. 595-604.

Kennedy, J. and Eberhart, R. (1995), "Particle swarm optimization", *IEEE International Conference on Neural Networks, Vol. 4*, pp. 1942-1948.

Khajemohammadi, H., Fanian, A. and Gulliver, T.A. (2014), "Efficient workflow scheduling for grid computing using a levelled multi-objective genetic algorithm", *Journal of Grid Computing*, Vol. 12 No. 4, pp. 637-663.

Kunzler, R., Muller-Quade, J. and Raub, D. (2009), "Secure computability of functions in the IT setting with dishonest majority and applications to long-term security", *TCC*, pp. 238-255.

Li, J. and Chinneck, J. (2009), "Performance model driven QoS guarantees and optimization in clouds", Software Engineering Challenges of Cloud Computing, CLOUD '09, ICSE Workshop, Ottawa, pp. 15-22.

Li, J., Lu, Q., Zhu, L., Bass, L., Xu, X., Sakr, S., Bannerman, P.L. and Liu, A. (2013), "Improving availability of cloud-based applications through deployment choices", *IEEE 6th International Conference on Cloud Computing*, pp. 43-51.

Liang, Y.-C., Tasgetiren, M.F., Sevkli, M. and Gencyilmaz, G. (2007), "A particle swarm optimization algorithm for makespan and total flowtime minimization in the permutation flowshop sequencing problem", *European Journal of Operational Research*, Vol. 177 No. 3, pp. 1930-1947.

Litoiu, M., Woodside, M., Wong, J., Ng, J. and Iszlai, G. (2010), "A business driven cloud optimization architecture", *Proceedings of 2010 ACM SAC*, pp. 380-385.

Murad, S. and Dowaji, S. (2014), "Value-Based Approach for SaaS Management", working paper, Damascus University, Damascus, 10 December.

Oriol, F. and Guitart, J. (2011), "Initial thoughts on business-driven IT management challenges in cloud computing providers", 6th IFIP/IEEE International Workshop on Business Driven IT Management, Dublin, OH, pp. 1070-1073.

Pandey, S., Wu, L., Guru, S. and Buyya, R. (2010), "A particle swarm optimization (PSO)-based heuristic for scheduling workflow applications in cloud computing environments", *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), Perth, 20-23 April.*

Pingale, S.H. and Mogal, V. (2015), "New approach to schedule Bag of Tasks (BoTs) with increase in parallel process", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3 No. 7.

Pretschner, A., Hilty, M., Schutz, F., Schaefer, C. and Walter, T. (2008), "Usage control enforcement: present and future", *IEEE Security & Privacy*, Vol. 6 No. 4, pp. 44-53.

Roy, A. and Dutta, D. (2013), "Dynamic load balancing: improve efficiency in cloud computing", *International Journal of Emerging Research in Management &Technology (IJERMT)*, Vol. 2 No. 4, pp. 22-29.

Sauvé, J., Marques, F., Moura, A., Sampaio, M., Jornada, J. and Radziuk, E. (2005), "SLA design from a business perspective", *Proceedings of the 16th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM, Citeseer*, p. 72.

Tsakalozos, K., Kllapi, H., Sitaridi, E., Roussopoulos, M., Paparas, D. and Delis, A. (2011), "Data engineering (ICDE)", *IEEE 27th International Conference on Data Engineering*, pp. 75-86.

Unruh, D. and Muller-Quade, J. (2010), "Universally composable incoercibility", *IACR Cryptology*, pp. 411-428.

Wu, K. (2014), "A tunable workflow scheduling algorithm based on particle swarm optimization for cloud computing", Master's projects, Paper No. 358, San José State University.

Yu, J. and Buyya, R. (2006), "A budget constrained scheduling of workflow applications on utility grids using genetic algorithms", Workshop on Workflows in Support of Large-Scale Science (WORKS), Paris, pp. 1-10.

Zhang, L., Chen, Y., Sun, R., Jing, S. and Yang, B. (2008), "A task scheduling algorithm based on PSO for grid computing", *International Journal of Computational Intelligence Research*, Vol. 4 No. 1, pp. 37-43.

About the authors

Salah Eddin Murad received his MSc in Software Engineering and Information Systems in 2012. He is certified Trainer and holds many certificates in project and risk management including PMP, PMI-RMP, IPMA. He occupied various positions including Head of PMO, Head of System Development Department, Trainer/Consultant and Engineering Instructor. Currently, he is a PhD Candidate in Software Engineering and Information Systems. His research interests include Cloud Computing, Software Engineering, Project and Risk Management, Agile Management, Business Driven IT Management, SOA, Semantic Web, System Integration and Software Architecture, BI and Big Data. Salah Eddin Murad is the corresponding author and can be contacted at: salah.eddin.murad@gmail.com

Salah Dowaji is an Associate Professor and the Dean of Information Technology Engineering Faculty at the Damascus University, where he has been since 2006. He obtained an Engineering Degree in Information Technology from the HIAST (Higher Institute of Applied Science and Technology), Damascus, Syria in 1989, and a Master Degree from the INPG (Institut National Polytechnique de Grenoble), Grenoble, France in 1991. He received his PhD. in Parallel Processing from the PARIS VI University (Université Pierre et Marie Curie – PARIS VI), Paris, France in 1995. From 1995 to 2006, he worked at the HIAST as a Teacher and a Researcher. His research interests centre on improving the understanding, design, and performance of parallel and networked computer systems, and software engineering. He is the author of many peer-reviewed publications.