# Towards Multi-Criteria Cloud Service Selection

Zia ur Rehman*, Farookh K. Hussain† and Omar K. Hussain‡
*Digital Ecosystem and Business Intelligence Institute*
*Curtin University, Perth, Australia.*
**zia.ur-rehman@curtin.edu.au*
†*farookh.hussain@cbs.curtin.edu.au*
‡*omar.hussain@cbs.curtin.edu.au*

***Abstract*—Cloud computing despite being in an early stage of adoption is becoming a popular choice for businesses to replace in-house IT infrastructure due to its technological advantages such as elastic computing and cost benefits resulting from pay-as-you-go pricing and economy of scale. These factors have led to a rapid increase in both the number of cloud vendors and services on offer. Given that cloud services could be characterized using multiple criteria (cost, pricing policy, performance etc.) it is important to have a methodology for selecting cloud services based on multiple criteria. Additionally, the end user requirements might map to different criteria of the cloud services. This diversity in services and the number of available options have complicated the process of service and vendor selection for prospective cloud users and there is a need for a comprehensive methodology for cloud service selection.**

**The existing research literature in cloud service selection is mostly concerned with comparison between similar services based on cost or performance benchmarks. In this paper we discuss and formalize the issue of cloud service selection in general and propose a multi-criteria cloud service selection methodology.**

***Keywords*-Multi-Criteria Decision Making; Cloud Computing; Service Selection**

## I. Introduction

Cloud computing is a new computing paradigm in which both hardware and software are provided to users over the Internet as services in the form of virtualized resources [1] with pay-as-you-go like pricing mechanisms [2].

Cloud services have been classified into several categories on the basis of various technical and economic aspects. The first classification based on service availability and includes two categories; public and private clouds. Public clouds have open access while access to private clouds is restricted to the owning organization and its subsidiaries. The second classification which includes; hardware as a service (HaaS), platform as a service (PaaS), software as a service (SaaS) and infrastructure as a service (IaaS), is based on type of service offered by the cloud [3]. In PaaS a cloud customer use the computing platform (e.g. Google's AppEngine) provided by a cloud to build its application. On the other hand IaaS cloud customers run their applications on cloud providers' virtual machines. This classification has been conceptualized in the form of a hierarchy by [4] and [5]. This conceptualization has practical significance in cloud computing because SaaS works on top of PaaS which itself depends upon IaaS. In general most smaller providers base their services on larger companies' infrastructure [6].

Furthermore, besides these differences in service models, pricing models also vary from provider to provider and from service to service on the same provider's infrastructure. Some (such as Google's AppEngine) charge users by the actual CPU cycles while others (such as Amazon ECC) charge in terms of actual virtual machine instances being utilized.

The increasing number of cloud providers, together with diverse type of services they offer on widely varying pricing schemes has lead to difficulties in comparing one cloud provider with another in terms of quality as well as cost of service which lead to complexities in cloud service selection. Furthermore cloud computing is dynamic due to the concept of virtual machine migration which has been made possible by virtualization which is one of the key cloud computing enabler technologies. Virtual machine migration allows seamless transfer of a running application from one virtual machine to another virtual machine which may be provided by a different IaaS provider.

Cloud computing is an attractive option for business due to its capabilities to provide better IT services with far less administrative overheads and technical complexities as compared with an in-house IT infrastructure at a much lower financial cost owing to the advantages of economy of scale [2] with added advantages of elastic computing. However the prime economical advantage that cloud computing has over in-house IT infrastructure is that due to the pay-as-you-go pricing policy the users only pay for computing resources which they actually use and do not need to pay for provisioning extra computing facilities which are only occasionally required to meet peak demands. Misra et al. [7] have done a comprehensive analysis to explore the viability of transition from conventional computing to cloud computing for various business entities and have concluded that a fairly large proportion of business do have financial benefits if they adapt cloud computing.

These are the key factors allowing both the cloud providers and the cloud users to generate revenues. The former by selling their unused computing resources and later by reduced expenditure on IT infrastructure. These advantageous aspects have led to a tremendous growth potential in cloud computing business. Leavitt [6] points out that cloud spending is expected to rise from about $16 billion in 2008 to $42 billion by 2012. Furthermore the same research also predicts that cloud computing will be 25% of annual IT expenditure growth by 2012 and can even become one third of the growth in 2013. The present economic crisis may encourage small companies to adopt cloud computing in order to cut costs [8]. This shows that service selection is a key issue in cloud computing and indicates that its importance will increase in future with expected expansion in cloud computing.

In this paper we propose a methodology for multi-criteria cloud service selection for selecting the service that best matches the cloud user's requirements from amongst numerous available services.

## II. Literature Review

Cloud service selection is a highly important research issue but it has not received much research attention and little literature has been published in this area because cloud computing itself is still in its early stages. In this section we give a brief overview of the related work in cloud service selection.

Gocinski et al. [9] have discussed the issue of cloud service discovery and selection in detail and have also discussed these issues in similar but not identical computing paradigms such as grid computing and clusters and have concluded that the techniques used in grid and cluster computing for service discovery and selection are still in infancy and have only seen marginal success even in their intended domains and are therefore unsuitable for cloud service selection.

One of the major obstacles in cloud service selection is due to the fact that the diversity in cloud computing offering makes it difficult to compare one cloud provider or service against others. To address this issue, Li A. et al. [10], [11] have proposed a service comparison methodology which is an important step towards vendor selection and have argued the importance of a comprehensive provider comparison framework. They have presented a framework for cloud service comparison called *"CloudCmp"* which is aimed at helping cloud users in selecting a cloud provider. CloudCmp consists of a set of benchmarking tools that are used to compare the common services (such as elastic computing cluster, persistent storage, intra-cloud and wide area network) and the benchmarking results are then used to predict the performance and cost of a cloud user's application when deployed on a cloud. An other important contribution is due to Han et al. [12]. They have proposed a conceptual framework for for a cloud service recommender system that relies comparison between available services on the basis of network QoS and virtual machine performance. An alternative method has been presented by Zeng et al. [4] using a maximum gain and minimum cost approach for optimal service selection. Godse and Malik [13] have argued that the problem of cloud service selection is an MCDM problem and have presented a case study of a sales force automation service using the Analytic Hierarchy Process (AHP) which is a well known techniques and is based on decomposing a complex MCDM problem into a system of hierarchies. Table one gives a summary of cloud service selection related literature.

Over the years, MCDM has emerged as an important new research area having immense practical significance in numerous scientific, logistic, engineering and industrial problems. It can be defined as a collection of methodologies for comparison, ranking and selecting multiple alternatives having multiple attributes [14]. MCDM methods, their classification and applications in various areas have been discussed in detail by [15] and [14].

Service selection, whether single or multi-criteria, falls with the preview of decision making since the end user has to make a decision to select a service from amongst available services. The typical properties of an MCDM problem as outlined by [15] and [16] are analogous to cloud service selection problem and underpin the notion for an MCDM based cloud service selection mechanism.

## III. Conceptual Framework

### A. Problem Formalization

An extremely important initial step in decision making is to formulate the decision problem into a formal and rigorous form [17]. In this section we present the cloud service selection problem in a generalized and abstract mathematical form and then build our method based on this mathematical model in the next subsection.

1) **Services (set):** Let $S = \{s_1, s_2, s_3, \dots s_l\}$ be a set of $l$ services where $l \geq 2$. This set contains all the service offerings from which a service is to be

Table I
AN OVERVIEW OF CLOUD SERVICE SELECTION LITERATURE

| Work | Theme | Approach |
|---|---|---|
| Gocinski et al. [9] | State of the art in cloud computing, Cloud service discovery and selection | Overview and comparison between cloud, grid and cluster computing |
| Li A. et al. [10] [11] | Cloud service comparison, application performance prediction and cost estimation. | Cloud benchmarks and testing of application prior before cloud deployment |
| Han et al. [12] | Cloud service recommender system | QoS and virtual machine performance of the available services are used as selection criteria |
| Zeng et al. [4] | Cloud service selection | Maximum gain and minimum cost optimization |
| Godse and Malik [13] | Cloud service selection | MCDM, Analytic Hierarchy Process (AHP) |

selected by the user (decision maker).

2) **Performance criteria (set):** Let $C = \{c_1, c_2, c_3, \ldots, c_m\}$ be the set of $m$ values where $m \geq 2$ and each $c_i \in C$ represents a criterion that may be a useful parameter for service selection.

3) **Performance measurement functions (set):** To each criteria $c_i \in C$ there corresponds a unique function $f_i \in F$ which when applied to a particular service, returns a value $p_i$ that is an assessment of its performance on a predefined scale. The set of all such functions may be defined as $F = \{f_1, f_2, f_3, \ldots, f_m\}$.

4) **Service descriptor (vector):** Let $D_i$ be a row vector ($1 \times n$ matrix) that describes a service $s_i \in S$ where each element $d_j$ of $D_i$ represents the performance or assessment of service $s_i$ under criteria $c_j \in C$. In other words $D_i = \begin{bmatrix} d_1 & d_2 & \ldots & d_m \end{bmatrix}$ where $d_j = f_j(s_i)$. Some of the criteria may be qualitative in nature and must be assigned quantitative values using the Likert-type scale [16]. Furthermore, these values must be normalized to eliminated computational problems resulting from dissimilarity in measurement units. The normalization procedure is used to obtain dimensionless units that are comparable. Consequently the larger the value becomes the more preference it has. The two most popular normalization methods are (1) linear normalization and (2) vector normalization [16].

5) **Decision matrix:** The service descriptor vectors $D_i$ can be combined to form the decision matrix A which is a $l \times n$ matrix :

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l,1} & a_{l,2} & \cdots & a_{l,n} \end{pmatrix}$$

Where each $a_{i,j}$ is the evaluation of service $i$ against criteria $j$ and is given by $a_{i,j} = f_k(s_k)$ where $c_i \in C$.

6) **User requirement criteria (vector):** Similar to our definition of service descriptor vector we define another vector $R = \begin{bmatrix} r_1 & r_2 & \ldots & r_m \end{bmatrix}$ where each $r_j$ is the user's (decision maker's) minimal requirement against criteria $c_j \in C$. These values must be scaled by the same technique previously used for service descriptor vector. The service selection process is fundamentally a comparison between the vector $D$ against all rows of the decision matrix followed by the selection of the serves whose description vector best matches with the user's requirement vector .

7) **User priority weights (vector):** Since individual users (decision maker) may have their own preferences therefore relative importance of a criteria may be different for each user. The user priority weights vector $W = \begin{bmatrix} w_1 & w_2 & \ldots & w_m \end{bmatrix}$ where each $w_i$ is the weight assigned by a user to criteria $c_i$ and value of 1 represents default weight.

The aim of service selection, which is subjective and depends on the relative importance given to each $p_i \in P$ by the user (i.e. decision maker), is to choose the best service from among all the services on offer. In order to simplify the problem we assume that cloud services remain unchanged during the decision process ( i.e. the values of

$\forall\ c_i \in C$ are constant). This assumption reduces into a decision problem without uncertainty.

### *B. Weighted Difference and Exponential Weighted Difference Service Selection.*

Our service selection process involves comparison between the user requirement criteria vector and all service descriptor vectors and then selection of the service which has the corresponding descriptor vector that best matches with the user requirement vector. This essentially involves a similarity measure between the user requirement vector and the service descriptor vectors. We perform these operations on decision matrix instead of individual descriptor vectors due to notational convenience and simplicity. There are three cases in this comparison i.e.

1) Descriptor vector exactly matches with the user requirement vector.
2) Descriptor vector has (generally) lower values then the requirement.
3) Descriptor vector has (generally) higher values then the requirement vector.

There are well established similarity measures in recommender system literature such as Pearson's Correlation, Cosine Similarity and Euclidean Distance. These similarity measurement methods find the similarity between two vectors, regardless of the difference being positive or negative and can not distinguish between a service ranked higher than user's requirement and a service ranked below the user's requirement. Therefore they produce useful results in the first two cases only and in the last case, where the offered service exceeds the user requirement they measure it as a dissimilarity. To circumvent this bottleneck we propose two different approaches.

In the first method, which we call Weighted Difference (WD), we subtract the user requirement vector form each row of the decision matrix D to obtain,

$$\begin{pmatrix} a_{1,1} - r_1 & a_{1,2} - r_2 & \dots & a_{1,n} - r_n \\ a_{2,1} - r_1 & a_{2,2} - r_2 & \dots & a_{2,n} - r_n \\ \vdots & \vdots & & \vdots \\ a_{l,1} - r_1 & a_{l,2} - r_2 & \dots & a_{l,n} - r_n \end{pmatrix}$$

In the next step we calculate the product of this matrix and transpose of the user requirement vector which yields this column vector,

$$\begin{pmatrix} (a_{1,1} - r_1)w_1 + (a_{1,2} - r_2)w_2 + \dots + (a_{1,n} - r_n)w_n \\ (a_{2,1} - r_1)w_1 + (a_{2,2} - r_2)w_2 + \dots + (a_{2,n} - r_n)w_n \\ \vdots \\ (a_{l,1} - r_1)w_1 + (a_{l,2} - r_2)w_2 + \dots + (a_{l,n} - r_n)w_n \end{pmatrix}$$

Each element of this column vector is an assessment of conformity of the respective service to the user's requirement and the element having the minimum value corresponds to the most appropriate service for the user.

This method, although computationally simple, does have one drastic shortcoming that during the calculation of ranking, the criteria where a service is below user requirement ($a_{1,1} < r$) get balanced by those exceeding user specification ($a_{1,1} > r$). We present another method to address this drawback.

In the second method which we call Exponential Weighted Difference (EWD), we make use of an exponential function to restrict the effect of mutual cancellation between criteria exceeding and below user requirement by multiplying the matrix by the scalar -1 and then replacing each element by $e$ raised to the power of the respective element itself i.e. replacing $a_{i,j}$ by $e^{-a_{i,j}}$ which yields the following matrix.

$$\begin{pmatrix} e^{-(a_{1,1} - r_1)w_1} + e^{-(a_{1,2} - r_2)w_2} + \dots + e^{-(a_{1,n} - r_n)w_n} \\ e^{-(a_{2,1} - r_1)w_1} + e^{-(a_{2,2} - r_2)w_2} + \dots + e^{-(a_{2,n} - r_n)w_n} \\ \vdots \\ e^{-(a_{l,1} - r_1)w_1} + e^{-(a_{l,2} - r_2)w_2} + \dots + e^{(a_{l,n} - r_n)w_n} \end{pmatrix}$$

In this approach the ability of a criterion to influence the similarity measure diminishes exponentially as it exceeds user requirement. Like the previous method each element of the resultant column vector is an assessment of the conformity of the corresponding service to the user's requirement and the element having minimum value corresponds to the most suitable service.

## IV. Conclusion

In this paper we have formalized the cloud service selection problem into a rigorous mathematical form and have presented a multi-criteria cloud service selection methodology by using this formalism.

In its present form our approach is only effective for service selection from amongst service offerings that are similar in specifications but only differ in performance. We have not taken into account several concepts such as reliability, trust and reputation which are very impotent in any service oriented environment such as cloud computing. We have also excluded the specifications of actual service criteria and practical implementation issues of our proposed performance evaluation functions from this paper all of which we intend to publish in a future work.

## Acknowledgment

This research has been funded by Curtin University under the Curtin International Postgraduate Research Scholarship (CIPRS) Program.

## References

[1] B. Furht, *Handbook of Cloud Computing*. Boston: Springer US, 2010, ch. Cloud Comp, pp. 3–19–19. [Online]. Available: http://www.springerlink.com/content/n767118183574r31/

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and Others, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?id=1721672

[3] B. P. Rimal, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems," *2009 Fifth International Joint Conference on INC, IMS and IDC*, pp. 44–51, Aug. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5331755

[4] W. Zeng, Y. Zhao, and J. Zeng, "Cloud service and service selection algorithm research," *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation - GEC '09*, p. 1045, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1543834.1544004

[5] W.-T. Tsai, X. Sun, and J. Balasooriya, "Service-Oriented Cloud Computing Architecture," *2010 Seventh International Conference on Information Technology: New Generations*, pp. 684–689, Apr. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5501650

[6] N. Leavitt, "Is cloud computing really ready for prime time?" *IT Professional*, vol. 27, no. January, p. 5, 2009. [Online]. Available: http://www2.computer.org/cms/Computer.org/ComputingNow/homepage/mostread/MostRead-CO-CloudComputingPrimeTime.pdf

[7] S. C. Misra and A. Mondal, "Identification of a companys suitability for the adoption of cloud computing and modelling its corresponding Return on Investment," *Mathematical and Computer Modelling*, vol. 53, no. 3-4, pp. 504–521, Mar. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S089571771000155X

[8] S. Qamar, N. Lal, and M. Singh, "Internet Ware Cloud Computing : Challenges," *(IJCSIS) International Journal of Computer Science and Information Security*, vol. 7, no. 3, pp. 206–210, 2010.

[9] A. Goscinski and M. Brock, "Toward dynamic and attribute based publication, discovery and selection for cloud computing," *Future Generation Computer Systems*, vol. 26, no. 7, pp. 947–970, 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167739X10000543

[10] A. Li, X. Yang, M. Zhang, and S. Kandula, "CloudCmp : Shopping for a Cloud Made Easy," in *2nd USENIX Workshop on Hot Topics in Cloud Computing*, 2010. [Online]. Available: http://www.usenix.org/events/hotcloud10/tech/full_papers/Li.pdf

[11] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: comparing public cloud providers," in *Proceedings of the 10th annual conference on Internet measurement*. ACM, 2010, pp. 1–14. [Online]. Available: http://portal.acm.org/citation.cfm?id=1879143

[12] S. M. Han, M. M. Hassan, C. W. Yoon, and E.-N. Huh, "Efficient service recommendation system for cloud computing market," *Proceedings of the 2nd International Conference on Interaction Sciences Information Technology, Culture and Human - ICIS '09*, pp. 839–845, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1655925.1656078

[13] M. Godse and S. Mulik, "An Approach for Selecting Software-as-a-Service ( SaaS ) Product," *Management*, pp. 155–158, 2009.

[14] U. Habiba and S. Asghar, "A survey on multi-criteria decision making approaches," in *2009 International Conference on Emerging Technologies*. Ieee, Oct. 2009, pp. 321–325. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5353151

[15] E. Triantaphyllou, B. Shu, S. Sanchez, and T. Ray, "Multi-criteria decision making: an operations research approach," *Encyclopedia of electrical and electronics engineering*, vol. 15, pp. 175–186, 1998.

[16] J. Lu, G. Zhang, D. Ruan, and F. Wu, *Multi-objective group decision making: methods, software and applications with fuzzy set techniques*. London: Imperial College Press, 2007.

[17] C. M. Brugha, "Structure of multi-criteria decision-making," *Journal of the Operational Research Society*, vol. 55, no. 11, pp. 1156–1168, Jun. 2004. [Online]. Available: http://www.palgrave-journals.com/doifinder/10.1057/palgrave.jors.2601777