On Selecting Composite Network-Cloud Services: A Quality-of-Service Based Approach

Minkailu Mohamed Jalloh
School of Communication & Information Engineering
Chongqing University of Posts & Telecommunications
Chongqing, 400065, P.R. China
minks2.jalloh@yahoo.com

Shitong Zhu, Fang Fang
School of Communication & Information Engineering
Chongqing University of Posts & Telecommunications
Chongqing, 400065, P.R. China
stzhu.ac@gmail.com

Jun Huang
School of Communication & Information Engineering
Chongqing University of Posts & Telecommunications
Chongqing, 400065, P.R. China
xiaoniuadmin@gmail.com

ABSTRACT

Cloud computing, by definition, refers to the on-demand delivery of shared computing resources via the Internet. As the number of both potential cloud service and network service providers grows, selecting a set of these services with superior performance has been an issue. In this paper, we first formulate the process of selecting the optimal composition from all functionally-equivalent services as a multi-criteria decision making problem. We then employ the Technique for Order Preference by Similarity to an Ideal Solution, a method acquiring the objective weight of each criteria in terms of its information entropy, to optimally select the composition of network-Cloud services. Our evaluation results suggest proposed scheme can achieve a close-to-optimal solution to implement in practice with considerable flexibility and simplicity.

CCS Concepts

•Networks → Network services; •Network services → Cloud computing;

Keywords

Cloud computing; Service selection; Service-Oriented Architecture; Multi-Criteria Decision Making; Quality-of-Service; TOPSIS

1. INTRODUCTION

In [10], cloud computing is defined as a computing model for enabling on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management or service provider interaction. It has five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. With this emerging paradigm, large upfront investments in hardware can be avoided and cloud providers easily deliver their computing services with pay-as-you-go pricing [1].

In the context of cloud computing, networking plays a pivotal role with the advent of network virtualization, for it decouples service provisioning from the network infrastructure and exposes underlying functionalities through resource abstraction. A recent study [8] on the performance of common commercial clouds indicates that the quality of networking poses significant impacts on the robustness of cloud computing. Therefore, non-functional, Quality-of-Service (QoS) factors including delay, pricing etc. are indeed crucial for selecting composite network-Cloud services. In the meantime, the problem of identifying the best candidate set from a group of functionally-equivalent services can be seen as a variant of multi-criteria decision making (MCDM) problem. However, in real-world situations, no single service can exceed all other services in terms of all criteria, but one may outperform in terms of several criteria while others are superior if judged on remaining criteria. Therefore, a trade-off among all criteria is necessary to be made for service selection to optimize overall performance [9].

MCDM is a branch of operations research that deals with decision making in which decisions are made according to a number of decision criteria rather than a single one. MCDM problems are mainly classified into two broad categories: 1) Multi-Criteria Selection Problems (MCSP) and 2) Multi-Criteria Mathematical Problems (MCMP) [11]. The former takes place when it is required to select the best candidate from a finite set of alternatives that are known a priori. The latter, instead, calls for scenarios where the size of alternative set is very large or even infinite and it is unknown a priori. In this paper, we focus on MCSP.

Some works have been addressed on this topic whereby most of them steer on either cloud computing or networking, not composite network-Cloud services. Han et al [5] compare available services based on network QoS, they highlight MCDM as a feasible formulation framework to service selection problems but do not carry it out. The TOPSIS is proposed by Yoon et al. [13] as a general solver to MCDM. It is in accordance with a derivation that an ideal candidate should have the shortest distance from the positive ideal solution (PIS) and farthest distance from the negative ideal solution (NIS) [2]. Huang et al. extend TOPSIS in combination with information entropy [7], form weighted Euclidean

RACS '15, October 09-12, 2015, Prague, Czech Republic

DOI: http://dx.doi.org/10.1145/2811411.2811546

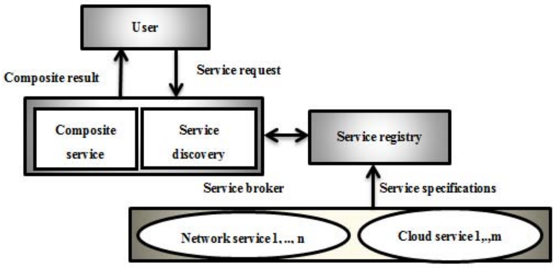

Figure 1: Network-Cloud convergence

distance as an intermediate metric, instead of the weighted decision matrix in original TOPSIS.

Network and Cloud convergence allows combined management control, and optimization of networking as well as computing resources in a cloud environment. Duan et al. explain [3] how QoS of Cloud services could be significantly improved when they are converged with high-quality network services. Figure 1 shows how both network and cloud service providers make their services available by publishing service descriptions at a service registry. Inside a complete workflow, a cloud customer firstly sends a request to the service broker that is in charge of searching the registry for the available services [4]. This request then travels through network services and eventually it is made available to the user as a composite service. An example could be Cloud user requesting his data to be processed using Amazon EC2 (Elastic Compute Cloud) in collaboration with Amazon S3 (Simple Storage Service) as storage space. In order to conduct requested services that are not necessarily located in the same geographical region, network services must be utilized through the Internet for data transmission, where Network-Cloud composition takes its place.

Suppose a biochemical factory generates 150 GB of data, and it processes and stores data in Cloud. If the factory has 10 virtual machines in one of the cloud providers and each of these machines has a processing time of 80 GB of data an hour, then the total cloud service time is 11.25 minutes. Given an implemented network with throughput for data transmission of 150 Mb/s to virtual machines, then even a single-trip transmission delay will be no less than one hour. Nevertheless it could be compressed to about 12 minutes if the throughput is otherwise at gigabyte measure. This indicates the importance of networking and the consequential necessity of network-Cloud service composition.

Recently, Service-Oriented Architecture (SOA), as an enabler of high-level encapsulation on networking and cloud computing from network-service providers, has attracted wide attention. These services can be published, discovered and finally selected and delivered to the user as a composite network-Cloud service as depicted in Figure 1 [8, 6].

One key challenge for selecting a composite network-Cloud service lies in the QoS-aware feature across networking and cloud computing domains. That is to say, those appropriate network-Cloud services must meet the end-to-end service requirement of the users. Applying the SOA design aspects is key to overcoming this difficulty by providing abstracted networking capabilities and exposing them to cloud computing systems. This new model is essentially a merger from two conventionally separate service components (networking and cloud computing) into one role from a perspective of providers. The concept of SOA through virtualization also helps reduce the complexity of network-Cloud composition [3].

The contribution made in this paper, as aforementioned, mainly marks a different consideration compared to preexisting literatures, for the scenario of selection amongst network-Cloud services in here underlines the composite property. It is a big enhancement in cloud computing architecture that is guaranteed to tame future research and industrial issues.

The reminder of this paper is organized as follows. Section 2 poses the well-formed problem formulation in terms of MCDM. Section 3 extensively examines proposed TOPSIS-based approach with corresponding algorithm. Section 4 shows numerical results of experiments. Lastly Section 5 casts concluding remarks.

2. PROBLEM FORMULATION

In this section, we present a well-formed problem formulation to assure the cogency and rigorousness of proposed approach.

Network-Cloud service selection in the form of a variant MCDM problem, can be expressed in the decision matrix D. Let$D = \{r_{ij}\}, 1 \le i \le m, 1 \le j \le n$, each row in D contains the numerical value (r_{ij}) representing the performance of a service against all criteria while each of the columns represent the performance of all services against one criterion. Similarly, r_{ij} is a measure of the performance of S_i under C_j. S is a finite set of services offered by the network-Cloud providers where $S = \{S_i\}, 1 \le i \le m$. Lastly, C is a finite set of the criteria on the basis of which services are selected. And $C = \{C_j\}, 1 \le j \le n$.

$$D = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_m \end{array} \begin{array}{c} \begin{array}{ccccc} \mathrm{C}_1 & \mathrm{C}_2 & \mathrm{C}_3 & \cdots & \mathrm{C}_\mathrm{n} \end{array} \\ \left(\begin{array}{ccccc} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ r_{31} & r_{32} & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \cdots & r_{mn} \end{array} \right) \end{array} \tag{1}$$

The attributes $C_1, \cdots, C_\mathrm{n}$ represent the memory, bandwidth, unit price and delay respectively. As the objective of this problem, it is required to select the optimal service composite out from $S_1, \cdots, S_\mathrm{m}$ through some satisfactory scheme.

3. PROPOSED APPROACH

In this section, we propose a TOPSIS-based approach to address the formed MCDM problem. Moreover, we examine the workflow of its corresponding algorithm in order to provide concrete tutorial in practice.

3.1 Weight Calculation

One of the most critical aspects of MCDM is to assess the unknown weights that act as importance indicators on different criteria. In our specific approach, the weights of attributes are obtained in accordance with information entropy, or Shannon entropy, a concept first introduced by Shannon [12]. The core idea behind the use of it is to quantify the homogeneity in certain dataset. More specifically, the attribute that maximizes the gain ratio or information gain the most is selected as the match candidate. For entropy is a measure of unpredictability of information content. Here we use the most general formula, the defining expression for entropy in the theory of information:

$$E_j = -K \sum\nolimits_{i=1}^{m} n_{ij} \ln n_{ij}, \tag{2}$$

where $j = 1, \cdots, n, K = \frac{1}{\ln m}$. The weights of the attributes C_j are

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{n} (1 - E_j)}. \tag{3}$$

Note that W is the weight set and $W = \{w_j\}$ where for all w_j, they satisfy

$$w_1 + w_2 + \cdots + w_n = 1. \tag{4}$$

3.2 Main Algorithm

To clarify the essence of proposed TOPSIS-based scheme, we divide the entire workflow of corresponding algorithm into seven separate but serializable steps:

Step 1: construct a decision matrix as in equation (1). Determine the m available services and the n available criteria.

Step 2: normalize the decision matrix D in order for the QoS values of different criteria to be comparable using the following equation

$$n_{ij} = \frac{r_{ij}}{\sqrt{\sum_{i=1}^{m} (r_{ij})^2}}. \tag{5}$$

With n_{ij}, the normalized decision matrix N can be formulated as follows

$$N = \begin{pmatrix} n_{11} & n_{12} & n_{13} & \cdots & n_{1n} \\ n_{21} & n_{22} & n_{23} & \cdots & n_{2n} \\ n_{31} & n_{32} & n_{33} & \cdots & n_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{m1} & n_{m2} & n_{m3} & \cdots & n_{mn} \end{pmatrix}. \tag{6}$$

Step 3: incorporate the weight matrix W supplied by Cloud's user body into normalized decision matrix N. It is worth noting that despite the use of entropy, pre-collection of user preference data via surveying or mining is otherwise satisfying to build W. Now the new multiplied weighted normalized decision matrix V is:

$$\begin{aligned} & V = N \cdot M \\ V &= \begin{pmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2n} \\ v_{31} & v_{32} & v_{33} & \cdots & v_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & v_{m3} & \cdots & v_{mn} \end{pmatrix} \\ &= \begin{pmatrix} n_{11} & n_{12} & n_{13} & \cdots & n_{1n} \\ n_{21} & n_{22} & n_{23} & \cdots & n_{2n} \\ n_{31} & n_{32} & n_{33} & \cdots & n_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{m1} & n_{m2} & n_{m3} & \cdots & n_{mn} \end{pmatrix} . \\ & \begin{pmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{pmatrix} . \end{aligned} \tag{7}$$

Step 4: determine the positive ideal solution (PIS) and the negative ideal solution (NIS).

$$\begin{aligned} A^+ &= \left\{ (\max v_{ij} | j \in J), \left(\min v_{ij} | j \in J^-\right) \right\} \\ &= \{v_1^+, \cdots, v_n^+\}, \end{aligned} \tag{8}$$

$$\begin{aligned} A^- &= \left\{ (\min v_{ij} | j \in J), \left(\max v_{ij} | j \in J^-\right) \right\} \\ &= \{v_1^-, \cdots, v_n^-\}, \end{aligned} \tag{9}$$

where $i = 1, 2, 3, \cdots, m$, A^+ and A^- are PIS and NIS, J and J^- are benefits and cost criteria respectively.

Step 5: compute separation measures using the N dimensional Euclidean distance. The separation measure D_i^+ of each service from the PIS is given as

$$D_i^+ = \left(\sum \left(v_{ij} - v_j^+\right)^2 \right)^{\frac{1}{2}}. \tag{10}$$

Similarly, the separation measure D_i^- of each service to the NIS can be computed as

$$D_i^- = \left(\sum \left(v_{ij} - v_j^-\right)^2 \right)^{\frac{1}{2}}. \tag{11}$$

Step 6: calculate the relative closeness to the PIS. The relative closeness to the service S_i with respect to the PIS A^+ is defined as

$$C_i^+ = D_i^- / \left(D_i^+ + D_i^-\right), \tag{12}$$

where $0 \leq C_i^+ \leq 1$ and $i = 1, \cdots, m$.

Step 7: lastly we sort and rank C_i^+ to see which one is the closest to the PIS.

4. NUMERICAL EXPERIMENTS

4.1 Experimental Setup

The experiment is setup using a laptop with 4GB Memory and 2.3 GHz of processor. In the meanwhile, MATLAB is utilized to input, trim and process data as well as to run numerical simulations. We randomly initialize the raw input to mimic the diversity of real network-Cloud service conditions.

4.2 Results

To track the complete workflow of selecting ideal composite network-Cloud service from a set of alternative services, Table 1 firstly shows the raw input data in which each criterion has its own scale.

Table 1: raw input service attribute data

Services	C_1	C_2	C_3	C_4	C_5
S_1	0.1689	0.131	0.1266	0.1333	0.1923
S_2	0.1858	0.1905	0.1519	0.16	0.1731
S_3	0.1926	0.1845	0.1646	0.1733	0.153
S_4	0.152	0.1667	0.1899	0.2	0.1538
S_5	0.1351	0.1488	0.1899	0.2	0.1346
S_6	0.1655	0.1786	0.1772	0.1333	0.1923

Table 2 shows the column-normalized matrix of raw service attributes.

In Table 3 and 4, it is observable that the smaller the entropy is, the higher its corresponding weight reaches. Such tendency suggests that more significance is attributed to the factor price than all other criteria, with the memory being the least influential.

Table 5 lists computed values of PIS and NIS.

Table 6 shows the relative closeness to the PIS. According to their values on C_i^+, services can be ranked as 6-2-3-1-4-5 in descending order. Therefore S_6 is said to be the selected service in our given raw input for its highest closeness.

Note that the designed system also has the flexibility to accept user-specified weights. For example, when it is users who supply weights 0.2000, 0.2000, 0.3000, 0.1500 and 0.1500 as $C_1 ... C_5$ respectively, the obtained results appear as 0.3302, 0.5805, 0.6699, 0.6036, 0.5494, and 0.6577 representing $S_1 ... S_6$ respectively. Given final closeness, S_3 that ranks the highest with 0.6699 ought to be selected as the best match.

5. CONCLUSION

Cloud computing is an emerging computing model which enables the on-demand access to shared computing resources. In the meantime, networking plays a pivotal role in cloud computing for it facilitates data communication among users and various cloud providers through the Internet. The growing number of cloud services around the world that have a wide ranging choices for potential cloud consumers has made network-Cloud service selection a big issue. In this paper, we first model such issue as a well-formed MCDM problem. Then we address it under a flexible framework that utilizes the information entropy theory to acquire the objective weights of criteria using a TOPSIS-based scheme. The numerical experiments show that the feasibility and practicality of proposed algorithm makes it applicable to real world situations.

Table 2: normalized input service attribute data

Services	C_1	C_2	C_3	C_4	C_5
S_1	0.1689	0.131	0.1266	0.1333	0.1923
S_2	0.1858	0.1905	0.1519	0.16	0.1731
S_3	0.1926	0.1845	0.1646	0.1733	0.1538
S_4	0.152	0.1667	0.1899	0.2	0.1538
S_5	0.1351	0.1488	0.1899	0.2	0.1346
S_6	0.1655	0.1786	0.1772	0.1333	0.1923

Table 3: entropy and weights

	C_1	C_2	C_3	C_4	C_5
Entropy	0.9962	0.9955	0.9948	0.9924	0.9954
Weight	0.1483	0.1752	0.2028	0.2964	0.1774

Table 4: normalized weights

Services	C_1	C_2	C_3	C_4	C_5
S_1	0.025	0.0229	0.0257	0.0395	0.0341
S_2	0.0275	0.0334	0.0308	0.0474	0.0307
S_3	0.0286	0.0323	0.0334	0.0514	0.0273
S_4	0.0225	0.0292	0.0385	0.0593	0.0273
S_5	0.02	0.0261	0.0385	0.0593	0.0239
S_6	0.0245	0.0313	0.0359	0.0395	0.0341

Table 5: PIS and NIS

	C_1	C_2	C_3	C_4	C_5
PIS	0.0286	0.0334	0.0385	0.0395	0.0239
NIS	0.02	0.0229	0.0257	0.0593	0.0341

Table 6: relative closeness

Services	D_i^+	D_i^-	C_i^+
S_1	0.0198	0.0204	0.5078
S_2	0.013	0.0185	0.5876
S_3	0.0134	0.0181	0.575
S_4	0.0213	0.016	0.4288
S_5	0.0227	0.0167	0.4238
S_6	0.0115	0.0242	0.6784

6. ACKNOWLEDGMENTS

This work is supported by NSFC under grant 61309031, Postdoctoral Science Foundation of China under grant 2014M551740, NSF of Chongqing under grant cstc2013jcyjA40026, S&T Research Program of Chongqing Municipal Education Commission under grant KJ130523, and CQUPT Research Fund for Young Scholars under grant A2012-79.

7. REFERENCES

[1] Armbrust, M., A. Fox, and R. Griffith. A view of cloud computing. *Communications of the Acm*, 53.4:50–58, 2010.

[2] e. a. Behzadian, Majid. A state-of the-art survey of topsis applications. *Expert Systems with Applications*, 39.17:13051ï£¡C13069, 2012.

[3] Q. Duan. Modeling and performance analysis on network virtualization for composite network-cloud service provisioning. *Services (SERVICES), 2011 IEEE World Congress on 2011*, pages 548–555, 2011.

[4] Duan, Qiang, Y. Yan, and A. V. Vasilakos. . A survey on service-oriented network virtualization toward convergence of networking and cloud computing. *Network and Service Management IEEE Transactions*, 9.4:373–392, 2012.

[5] e. a. Han, Seung Min. Efficient service recommendation system for cloud computing market. *Grid and Distributed Computing. Springer Berlin Heidelberg*, pages 117–124, 2009.

[6] e. a. Huang, Jun. Qos-aware service composition for converged network-cloud service provisioning. *2014 IEEE International Conference on Services Computing (SCC) IEEE Computer Society*, pages 67–74, 2014.

[7] J. Huang. Combining entropy weight and topsis method for information system selection. *Cybernetics and Intelligent Systems, 2008 IEEE Conference on IEEE*, pages 1281–1284, 2008.

[8] Huang, Jun, G. Liu, and Q. Duan. . On modeling and optimization for composite networkï£¡ccloud service provisioning. *Network and Computer Applications*, 45:35–43, 2014.

[9] O. HUSSAIN and F. HUSSAIN. Iaas cloud selection using mcdm methods. *In e-Business Engineering (ICEBE), 2012 IEEE Ninth International Conference on IEEE*, pages 246–251, 2012.

[10] Mell P. and Grance T. The nist definition of cloud computing. *Communications of the Acm*, 53.6:50–50, 2011.

[11] Ravindran, A. Ravi, and A. R. Ravindran. Operations research methodologies. *Crc Press*, 2008.

[12] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 27.3:3–55, 2001.

[13] K. Yoon. A reconciliation among discrete compromise solutions. *Journal of the Operational Research Society*, 38.3:277–286, 1987.