Considering Service Level Agreement to Improve Performance for Automatic Web Service Composition

Chatchawin Srivised[1], Saprangsit Mruetusatorn[2]

Faculty of Information Technology, Thai-Nichi Institute of Technology
1771/1, Pattanakarn Rd, Suan Luang, Bangkok, 10250, THAILAND
[1]sr.chatchawin_st@tni.ac.th
[2]saprangsit@tni.ac.th

***Abstract*— Autonomic service level agreement (SLA) research area nowadays is introducing concept for quality of service (QoS) that meets service consumer's requirements. The service consumer is able to choose multiple QoS levels or negotiate QoS and cost with service providers while execution to establish SLA automatically. Thus, it increase more opportunity for both a service consumer which is performs as a web service composition composer and a service provider to manage their own service efficiently. This paper designs the extended QoS service description by introducing a selected SLA template with multiple levels and the QoS of each level can be varied in specific time. Finally, we simulate results by employing the Genetic Algorithm to compose web services according to user's SLA constraints to show efficient SLA management which the composer performs as a service provider.**

***Keywords*— service composition, QoS, SLA, QoS model, service level management**

I. INTRODUCTION

In Service-oriented Architecture (SOA) environment, Quality of Service (QoS) is reflected to the service consumer satisfaction. Moreover Service Level Agreement (SLA) is a kind of electronic contract between a service provider and a service consumer to provide a service with a certain quality at a fixed point of time [1]. A different set of concrete web services may operate at different QoS measures. So we have to select an appropriate set of concrete web services that guarantees the fulfilment of a given SLA [2]. An SLA is in a management layer which is expressed on a high level between two partners [3]. Although SLA have been using for a long time. SLA has been still an early stage of SLA auto management research area on SOA.

D. Jianmin and Z. Zhou [1] has introduced an automatic SLA management life cycle including in web service composition environment. An automatic SLA management needs a QoS model in machine readable way to understand in both a service provider and a service consumer [4] in order to offer and request SLA with certain QoS and price for automatically negotiating and establishing an SLA contract. Moreover F. Wagner and et. al [5] has been introduced a novel of QoS model to concern possible dependencies on a time of execution or on an input data supplied to the service. These are the reasons why we should extend more complex QoS description to support the negotiation for a certain price and quality in the automatic web service composition manner to increase more flexibility of pricing model for a service like an online movie service. This paper proposes a conceptual QoS model which supports dependency on a time of execution, then employs the Genetic Algorithm from [2] called E3-MOGA. Thus, we can compose web services in a way of multiple levels of SLA named "Platinum", "Gold", "Silver", respectively by the higher quality and cost. Finally we show the performance improvement that attracts the service consumer to select a new style of QoS Model.

II. SERVICE COMPOSITION

To comply a dependency of an execution time for simulation evaluation, we have selected a web service composition workflow for an online movie compression service. This service has pricing model for days of the week, derived from a sales promotion model from Sunday through Saturday, which assume that Friday and Weekend tend to have a high number of customers watching an online movie service. On the other hand in weekdays tend to have a lower number of customers.

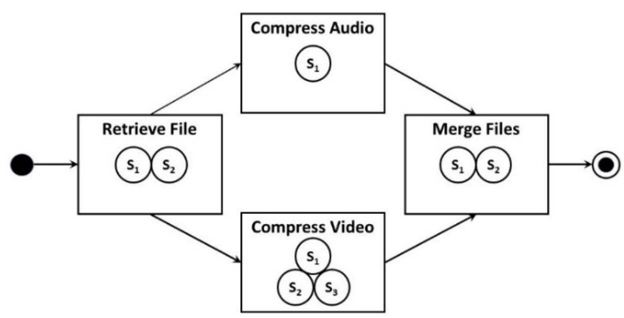

Fig. 1 An online movie compression service workflow [5]

Fig. 1 shows a workflow which is composed of 4 different abstract services named "retrieve file", "compress audio", "compress video", "merge files". Moreover, there are multiple service instances in the same function for some aforementioned services to be processed simultaneously.

III. QOS MODEL

A relationship between QoS and SLA is an aggregation from overall web service composition's QoS of each concrete service. In a QoS service description, a set of selected concrete web services is provided as one of an SLA level by overall QoS. In addition, a proposing QoS service description for multiple specific time, therefore, a set of selected concrete web services is able to vary in a specific time. Moreover, QoS in the conventional service description is able to transform into another kind of a management technique, for example, Key Performance Indicators (KPIs) [6]. The main equality between SLA and KPIs are service level objective (SLO) [7] which is a QoS term of quality together with probability. Therefore, the QoS model needs an extension to describe more complex quality metric like probability, indiscrete value and differentiate levels of quality to make QoS be able to vary in a specific time. In the other hand, SLO will be described at a higher description part in next step which to be guaranteed in term of an SLA obligation between a service consumer and a service provider. From all of the problems above, we employ an existing SLA description model from [8] to flexibly describe a quality value to solve them. Finally, Fig. 2 is a conceptual model which enables an automatic web service composition. And also allows the dependency on a time of execution.

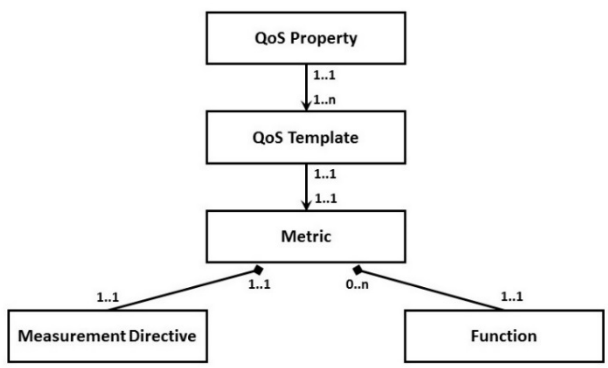

Fig. 2 Conceptual model to enable a dependency on a time of execution

QoS property describes a quality dimension such as throughput, latency and cost [9]-[12]. QoS template or service parameter represents a quality related to a metric for a particular state in a given time period. For example, cost of operation is 90 US dollar (a metric) between 00:00 and 23:59 on Saturday. A metric is a specification either how value is measured from a source by defining a measurement directive or how a metric is computed as defined in function. The function may also takes other Metrics and other input into account. For example, a resource metric measures operation cost is defined in measurement directive. A composite metric like "average number of invocations in a minute interval" is defined by the calculated average from a function specifying time series of reading of the resource metric together with invocations count from a measurement directive.

IV. MULTI-OBJECTIVE SLA OPTIMIZATION

Instead of aggregating the QoS values into a single unity value. E3-MOGA [2] is a multi-objective optimization a dominate relation which is a partial-ordering on the web service selections. Given an "individual" is a solution of web service composition represented by a gene set which is composed of 3 SLA level parts named "Platinum", "Gold" and "Silver". Each SLA level is a combination of genes which are dedicated for certain abstract services in composition workflow and also instance(s) of each concrete service. Fig. 3 shows an example of a gene set which is the one of an SLA level. Therefore, an individual is a sequence of 3 following gene sets for each SLA level.

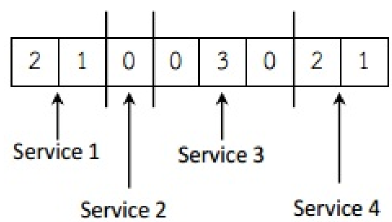

Fig. 3 An SLA level within an individual example

Fig. 4 is a pseudo code of E^3-MOGA which is maintain a set of individuals with higher fitness values, and then evolve their genes across generations through the use of genetic operation. (i.e., cross over and mutation) [2].

```
g = 0
P^0 = Randomly generated μ individuals, Q^0 = ∅
repeat until g = g_max {
  repeat until |Q^g| = μ {
    p1 = RWSelection(P^g), p2 = RWSelection(P^g)
    q  = Crossover(p1, p2)
    q  = Mutation(q)
    Q^g = Q^g ∪ q if q ∉ Q^g
  }
  P^(g+1) = Top μ of Sort(P^g ∪ Q^g)
  g=g+1
}

Crossover(p1, p2){
  for i = 1,...,n {
    center_i = (p1[i] + p2[i])/2
    q[i] = center_i + ((Fitness(p1) − Fitness(p2))|p1[i] − p2[i]|/2) / (F(p1) + F(p2))
  }
  return q
}
```

Fig. 4 E-MOGA's pseudo code [2]

TABLE I
VARIABLES AND FUNCTIONS IN E3-MOGA [2]

g_{max}	Maximum number of generations
μ	Population size
p^g	A set of individuals $\{p_k \mid 1 \leq k \leq \mu\}$ in elite population at g-th generation
p[i]	i-th gene of individual p, $(0 \leq i \leq n)$
Q^g	A set of offspring that is generated at g-th generation
Sort(P)	A function sorting P based on Fitness(p_k)
RWSelection(P)	A function performing a roulette wheel selection on P based on Fitness(p_k)
Fitness(P_k)	A function returning DRanking(p_k) / Density(p_k)
DRanking (P_k)	A function returning pk's domination rank
Density (P_k)	A function returning the density of p_k
Mutation (P_k)	A function randomly changing one of p_k's genes

To evaluate a fitness value for genetic operation, firstly, domination rank, indicates the excellence of an individual in the objective space. Let a set of objective values (QoS values from throughput, latency and cost) of an individual i be $\overrightarrow{o^i} = \left(\overrightarrow{o^i_{max}}, \overrightarrow{o^i_{min}}\right) = \left(o^{i,1}_{max}, \ldots, o^{i,n}_{max}, \ldots, o^{i,1}_{min}, \ldots, o^{i,n}_{min}\right)$ where $\overrightarrow{o^i_{max}}$ and o^i_{min} are sets of objective values to be maximized and minimized respectively. An individual i is said to dominate an individual j if both of the following conditions are true (1) $\{\forall \left(o^{i,k}_{max}, o^{j,k}_{max}\right)\ o^{i,k}_{max} \geq o^{j,k}_{max}\}$ and $\{\forall \left(o^{i,k}_{min}, o^{j,k}_{min}\right) \mid o^{i,k}_{min} \leq o^{j,k}_{min}\}$, all of QoS for each gene in i are better or equal to all of QoS for each gene in j, and (2) $\{\exists \left(o^{i,k}_{max}, o^{j,k}_{max}\right) \mid o^{i,k}_{max} > o^{j,k}_{max}\}$ or $\{\exists \left(o^{i,k}_{min}, o^{j,k}_{min}\right) \mid o^{i,k}_{min} < o^{j,k}_{min}\}$, at least one of QoS for each gene in i is better than QoS for each gene in j.

Secondly, constraint-dominate considers constraint of objectives with a degree of violation. For example a constraint specifies throughput must be over 40,000 byte per second. When an individual cannot satisfy a constraint, the individual is said to be infeasible. An individual i is said to constraint-dominate an individual j if any the following conditions are true (1) An individual i is feasible and individual j is not, (2) both i and j are feasible and i dominates j with respect to their objective values, (3) both i and j are infeasible but i dominates j with respect to their constraint violations, which is the less violations one.

Thirdly, to evaluate dominate rank for each individual, count a number of the constraint-dominate compare to each other for every individuals in particular generation. Next, arrange a number of constraint-dominate in ascending order. The domination rank is calculated by μ - $ndom_i$ (a number of constraint-dominate above). Therefore, an individual which has a lower number of constraint-dominate to other individuals will have a higher domination rank.

Last step, a density indicates the diversity of individuals in the objective space. The objective space is divided into certain size of K-dimensional cells (K is the number of objectives). The number of individuals in each cell is defined as the density of the cell, and the density of an individual is equal to the density of the cell where is the individual is located. Therefore, individuals with lower densities will obtain a higher fitness to improve the density of individuals.

V. SIMULATION EVALUATION

We apply a simulation environment to evaluate performance improvement of the proposed QoS model. First, we have to create both of the conventional and the proposed QoS service description. The conventional QoS values is selected from service description in [2] by obtaining QoS values from the highest probability. And the proposed QoS values, Cost values are derived from a sales promotion model for each days of the week. Fig. 5 shows cost percent rates which to be used to convert the conventional cost to create the proposed QoS service description. On the other hand, throughput and latency are remain the same.

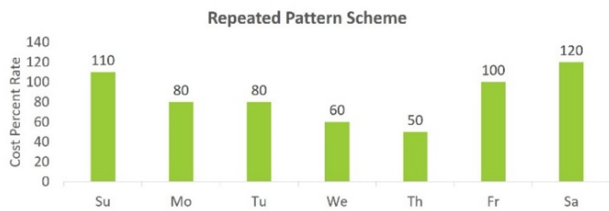

Fig. 5 Cost percent rate derived from the conventional

Hence, the proposed cost values on Sunday are higher than the conventional by 110 percent according to a cost percent rate above, cost values on Monday are lower than the conventional by 80 percent and so on.

A. Simulation Configurations

Composition constraints for multiple SLA levels offered to a service consumer have to be created. A service provider either specifies every constraint values clearly or leaves some wildcard for an automatic composition to freely select web services within a total cost to cover the budget.

TABLE II
COMPOSITION CONSTRAINTS

Service Level	Constraint			
	Through put (B/s)	Latency (ms)	Cost (US)	Total Cost
Platinum	40,000	550		2,000
Gold	6,000	450		
Silver	2,000		250	

Table II is constraints for simulation evaluation in every composition simulation. A Platinum level is the most expensive and also the highest quality for a service consumer. Where throughput is a QoS attribute which is to be maximized, in contrast, latency and cost are to be minimized.

For genetic algorithm parameters, the maximum number of generations was set to 500 and 100 for the population size

B. Simulation Results

The results are divided into 2 groups (1) aggregated QoS (2) selected concrete web services compare between the conventional and the proposed QoS model. The aggregated QoS shows quality on weekdays tend to gain the higher performance.

Fig. 6 Selected throughput from the proposed QoS model

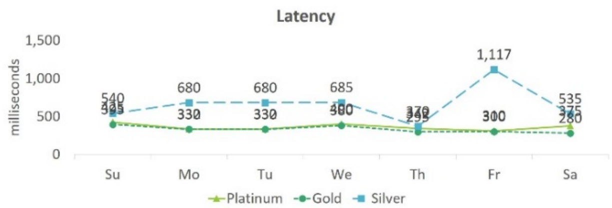

Fig. 7 Selected latency from the proposed QoS Model

Fig. 8 Selected cost from the proposed QoS model

By assumption that quality on Friday is a selection of the conventional service description. Selections on weekdays tend to be higher than weekend. In contrast, on weekend tends to be lower. These selections are applied to all service levels even on some day in Gold and Silver are not obviously separate the quality level like in Platinum.

The selected concrete web services compare between the conventional and the proposed QoS model, Selections of the conventional tend to lower than the proposed. Result in Table III lists numbers of selected concrete web services from Sunday through Saturday for every service levels.

In a Gold level shows equal number of selection. The other levels show numbers of the proposed QoS model service selection is higher than the conventional one. Especially in a Silver level.

Fig. 9 Selected concrete web services for Platinum level

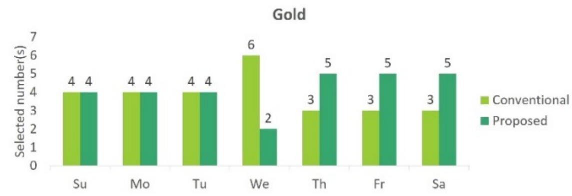

Fig. 10 Selected concrete web services for Gold level

Fig. 11 Selected concrete web services for Silver level

Fig. 9-11 show some of selections not obviously select only the proposed QoS model. But overall of the selections tend to select the proposed QoS model.

VI. CONCLUSIONS

This paper designs QoS model for web service description to conform a trend of the automatic SLA management research. Furthermore the model is used for describe QoS variation in a specific time. The simulation's results show the performance improvement on the sales promotion time. Moreover, service descriptions which are using the proposed QoS model are likely to be selected more than using the conventional one. Hence, the proposed QoS model absolutely attracts a service consumer to select a particular service if it is exploiting a sales promotion pricing model and also tend to gain higher performance or even reduce cost for an automatic web service composition.

TABLE III
SELECTED CONCRETE WEB SERVICES COMPARED BETWEEN THE CONVENTIONAL AND THE PROPOSED QOS MODEL

Service Level	Sunday		Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Total		Select Proposed
	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	C.	P.	
Platinum	4	4	3	5	3	5	5	3	3	5	4	4	4	4	26	30	4
Gold	4	4	4	4	4	4	6	2	3	5	3	5	3	5	27	29	2
Silver	3	5	3	5	3	5	3	5	1	7	2	6	3	5	18	38	20

REFERENCES

[1] D. Jianmin and Z. Zhuo, “Towards Autonomic SLA Management: A Review,” in *IEEE International Conference on Systems and Informatic*, Yantai, CH, May 19-20, 2012, pp. 2552-2555.

[2] H. Wada and et al, “Multiobjective Optimization of SLA-aware Service Composition,” in *IEEE Congress on Services 2008 - Part I*, Honolulu, HI, USA, July 6-11, 2008, pp. 368-375, 2008.

[3] Y. Leng, "Semantics-Aware Planning Methodology For Automatic Web Service Composition," Ph.D. dissertation (Mathematics and Natural Sciences), University of Bonn, Bonn, DE, 2012.

[4] P. Bianco, G. A. Lewis and P. Merson, *Service Level Agreements in Service-Oriented Architecture Environment,* Carnegie Mellon University: Research Showcase (Department of Software Engineering Institute), 2008.

[5] F. Wagner and et al, “Multi-objective Service Composition with Time- and Input-Dependent QoS,” in *IEEE International Conference on Web Services 19th*, Honolulu, HI, USA, June 24-29, 2012, pp. 234-241.

[6] S. A. Aloussi, “SLA Business Management Based on Key Performance Indicators,” *Proceedings of the World Congress on Engineering, vol. 3,* London, UK, July 4-16, 2012, pp. no page.

[7] A. Andrieux and et al, *Web Services Agreement Specification (WS-Agreement),* Grid Resource Allocation Agreement Protocol Working Group, 2007

[8] H. Ludwig and et al, *Web Service Level Agreement (WSLA) Language Specification,* IBM, 2003.

[9] M. O. Hilary, “Quality of Service (QoS) in SOA Systems,” M.S. thesis, Dept. Computer Languages and Systems, Polytechnic University of Catalonia, Catalonia, ES, 2009.

[10] S. Sounsri,“ An Automated Web Services Composition using Genetic Algorithm,” M.S. thesis, Dept. Science, Kasetsart University, Bangkok, 2004.

[11] W. A. Rahman, W. Nurhayati and F. Meziane, “Challenges to Describe QoS Requirements for Web Services Quality Prediction to Support Web Services Interoperability in Electronic Commerce,” *Communications of the IBIMA, vol. 4,* pp. 50-58, 2008.

[12] T. Wasino, “Using OWL-S Process Model with Constraints on Functional Behavior for Automatic Web Services Composition,” M.S. thesis, Dept. Engineering, Chulalongkorn University, Bangkok, 2007