Improving Consumer Satisfaction through Building an Allocation Cloud

Kuo Chen Wu
Institute of Computer and Communication Engineering
National Cheng Kung University
Tainan, Taiwan
kuchenwu@nature.ee.ncku.edu.tw

Hewijin Christine Jiau,   Kuo-Feng Ssu
Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
{jiauhjc, ssu}@mail.ncku.edu.tw

Abstract—**Consumer satisfaction depends on the quality of service. Services with high quality usually lead to high cost. Due to the budget limitation, consumers tend to acquire services with lowest cost. Feasible request allocation can improve service quality or reduce service cost. Some request allocation algorithms need the coordination with service consumers, but the benefits from consumers' coordination are arranged by service providers, especially for on-demand consumers. On-demand consumers have no idea to improve service quality or reduce service cost through allocating their requests, even though service providers can collect on-demand requests to perform feasible request allocation for efficient service execution. To solve this issue, an allocation cloud is proposed in this study. The allocation cloud assists consumers in allocating requests to optimize consumer satisfaction. Because the allocation cloud contains various allocation algorithms and collects a large amount of service requests, the allocation cloud can coordinate service consumers to follow feasible request allocation algorithms. Moreover, the allocation cloud can reveal the relation between service consumers' coordination and the improvement of service execution. The revealed relation can help service consumers decide beneficial strategies to improve their satisfaction. To evaluate the effectiveness of our approach, a study of video delivery service is conducted. The results confirm that consumer satisfaction can be improved through the proposed allocation cloud.**

Keywords-Cloud Computing; Quality of Service; Request Allocation; Consumer Satisfaction.

I. INTRODUCTION

High quality of service (QoS) leads to high consumer satisfaction and also means high service cost. For example, service respond time increasing usually needs to increase network bandwidth, computation power, or cache size. The addition of resources means the cost increasing. However, service consumers always has budget limitation and need to consider the trade-off between QoS improvement and service cost increasing. Besides considering the trade-off between QoS and service cost, consumers can try to reduce through two approaches. One approach is making a favourable service agreement with service providers and the other approach is performing service sharing through appropriate request allocation. However, On-Demand consumers ask services when they need and usually send small numbers of requests. Their characteristics make the cost reduction

through request allocation or service agreement becomes difficult. The reasons are listed as follows:

1) The amount of service requests from each On-Demand consumer is small. It is hard for on-demand consumers to gain benefits by performing request allocation with a small amount of requests.
2) An on-demand consumer cannot guarantee the amount of service requests. Therefore, service providers do not make a favourable service agreement with an On-Demand consumer.

Therefore, it is difficult for On-Demand consumers alone to improve consumer satisfaction through request allocation and service agreements. However, via collecting the requests from on-demand consumers, service providers can still perform suitable request allocation for efficient service execution. Through the suitable request allocation, on-demand consumers can acquire efficient service execution. However, on-demand consumers have no idea to coordinate with other on-demand consumers to follow feasible requestion allocation and gain additional benefits.

We propose an allocation cloud to coordinate on-demand consumers for efficient request allocation. The allocation cloud contains two important concepts. The first concept is the collection of on-demand consumers' requests. The allocation cloud coordinates on-demand consumers to follow request allocation according to these on-demand consumers' requests. The requests contain the functional and non-functional requirements of requesting services. The non-functional requirements are presented as hard constraints and soft constraints [1] in this study. Hard and soft constraints are criteria to evaluate request allocation in this study. Furthermore, on-demand consumers' coordination and request allocation are based on the evaluation of hard and soft constraints. The second concept is the request allocation algorithm set. A request allocation algorithm specifies targeted request collection, performs request allocation for the request collection, and presents the effects of applying the request allocation. Multiple allocation algorithms are available in the allocation cloud for different request collections. With these allocation algorithms, the allocation cloud can coordinate consumers to follow feasible allocation

algorithms for corresponding request collection. In addition, the allocation cloud can further improve service execution using how consumers' coordination for more efficient resource usage or better QoS. Keeping the relation between consumers' coordination and service execution is important for making service agreements with service providers. On-demand consumers and the allocation cloud can be aware that their coordination really contributes the service quality improvement. Therefore, the allocation cloud makes service agreements with service providers for improving QoS or reducing prices.

This paper is organized as follows: Section II lists relevant studies which discuss the relation between consumer satisfaction and cloud services. Section III introduces the major concept of the allocation cloud. Section IV discusses the effectiveness of the allocation cloud through an experiment. The experiment which simulates the video delivery service with *Coordinated Channel Allocation* [2] (COCA) is attached. Section V is the conclusion of this paper.

II. RELATED WORK

Due to increasing popularity of cloud computing [3][4], more and more service providers utilize clouds as their deployment platforms [5][6][7]. There are sufficient resources for service executing in cloud platforms, and how to allocate service requests adequately and optimize profit [8] or consumer satisfaction becomes important.

Several studies define and explore consumer satisfaction of using cloud services. Tsakalozos et al. [9] explore the consideration of *Infrastructure as a Service* (*IaaS*) cloud administrator to maximize cloud-consumer satisfaction. They map cloud-consumer satisfaction into financial profit, which is defined as the difference between revenue and cost. Both of revenue and cost are impacted by the cloud's physical resource usage. Besides that, there are two assumptions in this study. The first assumption is that a cloud-consumer has a budget constraint. The second one is that the cloud's physical resources are limited. They develop the approach for cloud administrators to maximize per-consumer financial profit through allocating physical resources.

Kantere et al. [10] propose an pricing method to optimize the profit of cloud cache services. They expect that consumer dissatisfaction from high service charge leads to the dropping of service demand. Besides that, the cloud risks to permanently lose the dissatisfied consumers. Base on this expectation, consumer satisfaction is the factor of optimizing profit. The consumer satisfaction is an altruistic tend of the optimization that is opposite to the egoistic tend of cloud profit. They express altruistic tend as: 1) a guarantee for a low limit on consumer satisfaction, or 2) an additional maximization objective. According to the expression, they optimize cloud cache service profit with consider consumer satisfaction.
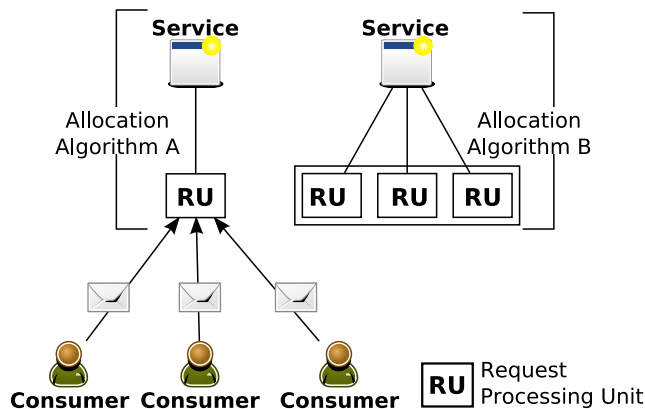

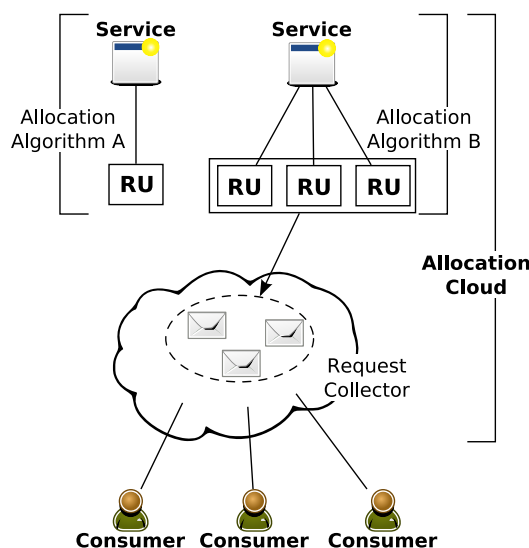Figure 1.   Request allocation without the allocation cloud.

Figure 2.   Request allocation within the allocation cloud.

Chen et al. [11] develop an utility model for measuring consumer satisfaction in the cloud. Since they focus on the service provisioning problem, the utility model defines consumer satisfaction with two factors: service price and response time. They assume that consumer satisfaction is decreased with higher service price and longer response time. The developed utility model can help service providers build services in the *IaaS* cloud platform with maximal profit.

III. ALLOCATION CLOUD

Consumers want to perform efficient request allocation to share services and reduce cost. However, consumers are limited to perform efficient request allocation because they ask for services independently and usually have a small amount of requests. Figure 1 illustrates such situation. In Figure 1, each consumer has one request to the same service. The desired service provides two allocation algorithms. Both the algorithm A and the algorithm B reserve request processing

units to serve request. The algorithm B allows three the same requests to share the service. However, consumers can only perform request allocation with allocation algorithm A since each consumer does not know they can cooperate to perform allocation algorithm B. The allocation cloud is proposed to solve this issue. Figure 2 illustrates the overview of the allocation cloud. In Figure 2, the allocation cloud includes several allocation algorithms and a request collector. The allocation cloud collects consumers' requests and reserves them in the request collector. The allocation cloud can find several requests that can be shared with the same service because the allocation cloud can evaluate the global requesting situation. The allocation cloud subsequently performs request allocation with allocation algorithm B. As a result of service sharing, consumers can reduce cost and improve satisfaction through the allocation cloud.

Besides the request allocation, there are two operations in the allocation cloud which can improve consumer satisfaction. The first operation is the service agreement through the allocation cloud. Because the allocation cloud reserves requests, the allocation cloud can predict service requesting situation and make service agreements according to the prediction. While the actual service requesting fulfils the prediction, the cost can be reduced and consumer satisfaction can be improved. The second operation is the consumer negotiation in the allocation cloud. The allocation cloud can negotiate service requesting with consumers when actual service requesting is not conformed to the expected situation. If consumers accept to modify their requesting to fulfil the expectation, consumer satisfaction can be improved.

In the following sections, the definition of consumer satisfaction is described in Section III-A. Operations of the allocation cloud, request allocation, service agreement, and consumer negotiation, are explained in Section III-B, III-C and III-D, respectively. Finally, Section III-E discusses the restriction of building an allocation cloud.

A. Consumer Satisfaction

Consumer satisfaction in the allocation cloud is the sum of all request satisfaction. Given all requests $R = \{r_1, \ldots, r_m\}$ in the cloud, every request satisfaction is impacted by two factors. One factor is the evaluation on the quality q_r that the service exhibits and the other factor is the service price. QE_r is the function of quality evaluation for request r and is specified by hard constraints, C_{Hr}, and soft constraints, C_{Sr}. The quality evaluation QE_r on q_r is presented as follows.

$$QE_r(q_r) = C_{Hr}(q_r) \cdot C_{Sr}(q_r), where \quad (1)$$

$$\begin{cases} C_{Hr}(q_r) & \in & \{0, 1\} \\ C_{Sr}(q_r) & = & [0.0, 1.0] \end{cases}$$

If q_r does not conform to C_{Hr}, the value of $C_{Hr}(q_r)$ is 0 that means that q_r is not satisfied with hard constraints. Therefore, q_r must conform to C_{Hr} first and the influence of

$C_{Sr}(q_r)$ is revealed. Based on QE_r, the request satisfaction S_r is presented as follows.

$$S_r(q_r) = \frac{QE_r(q_r)}{P} \quad (2)$$

The service price is represented as P. The low price and the high quality of service lead the high request satisfaction. Based on S_r, consumer satisfaction CS in the allocation cloud is presented as follows.

$$CS = \frac{\sum_{r \in R} S_r(q_r)}{\mid R \mid} \quad (3)$$

The CS function is defined as the ratio of total request satisfaction values to the total request amount in the allocation cloud. It should be noted that q_r and P exhibited by a service vary upon different service types. Also, the specifications of C_{Hr} and C_{Sr} are related to service requests. Based on CS function, the allocation cloud can improve consumer satisfaction by performing optimization techniques. The cloud operations on request allocation, service agreement, and consumer negotiation can be executed to achieve the optimization.

B. Request Allocation

The allocation cloud can perform alternative allocation algorithms. Different allocation algorithms lead to a change of service quality and service cost. The allocation cloud needs to estimate the change of q_r and P and attempts to optimize consumer satisfaction defined in (3) from available allocation algorithms. Besides, there are some rules that should be fulfilled before allocating requests, as listed as follows.

$$\forall r \in R, C_{Hr}(q_r) \leq C_{Hr}(q'_r) \quad (4)$$

$$\sum_{r \in R} QE_r(q_r) \leq \sum_{r \in R} QE_r(q'_r) \quad (5)$$

In (4) and (5), q'_r is the service quality with alternative allocation algorithms. Equation (4) ensures that q'_r cannot be worse that q_r. If q_r conforms to C_{Hr}, q'_r must also conform to C_{Hr}. The allocation cloud needs to make sure that all involved consumers are not sacrificed in alternative allocation algorithms. Equation (5) ensures that $\sum_{r \in R} QE_r(q_r)$ cannot be worse within alternative allocation algorithms. The allocation cloud needs to promise that QoS of all requests are not sacrificed in alternative allocation algorithms.

C. Service Agreement

Request allocation in the allocation cloud can lead request fulfilment with lowest cost under the prerequisite of (4) and (5). Besides, the allocation cloud has to prepare enough services for requests with expected quality constraints. The allocation cloud, instead of consumers, performs service agreements with service providers to prepare enough services. Figure 3 illustrates the communication protocol for
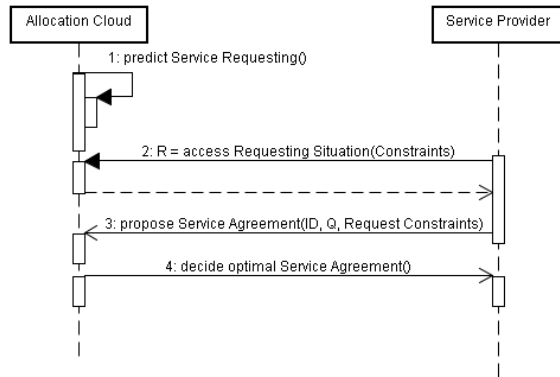
Figure 3. A service agreement protocol.

making service agreements between the allocation cloud and service providers. As shown in Figure 3, the allocation cloud does not find suitable services through service brokers and does not actively contact with service providers. Alternatively, service providers send query messages to the allocation cloud to check the service requesting situation. After confirming the service requesting situation, service providers can propose suitable service agreements according to this situation.

The communication protocol has four steps, as follows

1) *predict Service Requesting():* The allocation cloud can collect historical service requesting records. Therefore, the allocation cloud predicts service requesting according to historical service requesting records and all other related information. The prediction is represented as $R = \{r_1, \ldots, r_m\}$ and $\forall r \in R$ includes C_{Hr} and C_{Sr}

2) *access Requesting Situation(Query Statement):* Service providers query service requesting situation in this step. Service providers can specify query messages, such as needed service amount or occupied time. Service providers can concentrate on target service requesting through specifying query messages. The allocation cloud receives the query messages and returns service requesting situation according to the query messages.

3) *propose Service Agreement(ID, Q, Request Constraints):* Service providers can decide whether to propose service agreements or not according to the received service requesting situation. If service providers judge that they can propose profitable service agreements, they continue to propose service agreements to the allocation cloud. The proposed service agreements need to contain three parameters, ID, Q, and Request Constraints. The ID is the identification of this service agreement and is used by the allocation cloud to recognize this service agreement. The Q is QoS for this service agreement. The allocation cloud uses this parameter to judge consumer satisfaction of this service agreement. The Request Constraints are the constraints which the allocated

requests must be conformed to. Service providers use this parameter to specify target service requesting.

4) *decide optimal Service Agreements():* The purpose of step 4 is to select a set of service agreements from all received service agreements. This selected set of service agreements can optimize consumer satisfaction in the allocation cloud. The fitness function of the optimization is the CS function defined in (3). The allocation cloud simulates the request allocation with different allocation algorithms and calculates the value of CS. The allocation cloud returns selected service providers to achieve service agreements when deciding the set of service agreements which can optimize consumer satisfaction.

This communication protocol is proposed for the listed reasons.

- The allocation cloud can trigger service providers to propose service agreements actively because the allocation cloud keeps a large amount of service requests.
- Service providers can be aware of the service requesting situation through the allocation cloud. Service providers can propose the most profitable service agreements specialized for target service requesting.
- The allocation cloud does not need to specify service specification and find suitable services. The allocation cloud can concentrate on comparing proposed service agreements and on deciding the set of service agreements which lead to consumer satisfaction optimization.
- Because service providers only propose their service agreements for target service requesting, the amount of proposed service agreements is less than the amount of all service agreements which service providers can provide. This situation reduces the complexity of achieving service agreements

D. Consumer Negotiation

The allocation cloud makes service agreements based on the expected service requesting situation. Consumer satisfaction optimization is achieved in accordance with the expected situation. However, actual service requesting situation is different with the expected situation and consumer satisfaction is not as good as the expectation. Consumer satisfaction can be improved when the actual service requests are adjusted to fulfil the expected situation. For this purpose, the allocation cloud negotiates quality constraint refinement of service requests with consumers. The allocation cloud negotiates with consumers who send requests fulfilled the conditions, as listed as follows.

- The service request $r \in R$ proposed by a consumer is similar to the expected service request for a specific service.
- Request satisfaction S_r will be improved if the allocation cloud modifies quality constraints of this request r.

- Consumer satisfaction CS in the allocation cloud will also be improved if the allocation cloud modifies quality constraints of this request r.

The purpose of the allocation cloud is to serve consumers and improve consumer satisfaction. Therefore, the allocation cloud must allow consumers to be aware of the modification of quality constraints and request satisfaction.

E. Restrictions

The major restriction of building an allocation cloud are privacy and scalability concerns. The administrator of the allocation cloud must be trusted by target consumers because the allocation cloud collects service requests from consumers. Besides that, the allocation cloud must collect enough requests in order to trigger service providers to propose their service agreements for the allocation cloud.

IV. EXPERIMENT

An experiment is conducted to evaluate the effectiveness of the allocation cloud. In this experiment, streaming video delivery service is selected as the targeted application. Due to the high resource need and heavy workload, selectively feasible allocation algorithm is important to guarantee system capability with efficient resource utilization. Some allocation algorithms need to coordinate with service consumers, such as *coordinated channel allocation* [2] (explained in the next section). However, the benefits of request allocation algorithms only reflects on the service providers, even though the request allocation algorithms need consumers' coordination. We regard these allocation algorithms as services in the allocation cloud for service consumers and evaluate the improvement of consumer satisfaction using the allocation cloud. There are two assumptions for resolving the concerns for the restrictions mentioned in Section III-E. The first assumption is that consumers of the video delivery service have minor privacy concerns and allow the allocation cloud to allocate their video requests. The second assumption is that the allocation cloud can collect large amount of requests. We make the second assumption because streaming video is popular and most consumers request video on-demand. In the following contents, coordinated channel allocation is introduced. The benefits of the allocation cloud is explained while applying COCA to the allocation cloud.

A. Coordinated Channel Allocation for Video Delivery Services

COCA is based on batching techniques [12][13][14]. Therefore, these requests are grouped together in a small interval. A scheduler with COCA negotiates with consumers for reserving the requests with larger resource needs and sets a deadline of starting service for each reserved group. Based on the reserved groups, the scheduler with COCA has two queues, reservation queue and regular queue. The reservation queue keeps all reserving groups and performs sort on groups according to group deadline. The regular queue keeps all awaiting groups and performs sort on all groups according to group arrival time and resource needs. When there exists free channels and the deadline of the group in the reservation queue is reached, the scheduler with COCA allocate the group in the reservation queue to a free channel. When there exists free channels but no group deadline is reached, the Scheduler with COCA allocate the group in the regular queue to a free channel. This algorithm avoids that the groups with large resource needs occupy channels for a long time and hinders other groups from using the service.

According to COCA, delivery service providers perform efficient request allocation based on the arrival requests and the coordination with consumers. In other words, if consumers can coordinate with each other before requesting services, consumers can negotiate with service providers for additional benefits. However, there are two reasons make on-demand consumers unable to coordinate with each other. First, on-demand consumers do not know other arrival requests. Since consumers have no idea of other arrival requests, they cannot coordinate with each other. Second, on-demand consumers have no idea about efficient request allocations for the video delivery service. Moreover, video on-demand consumers are usually end users and do not have background knowledge of efficient request allocations. Even if they know other arrival requests, they still cannot coordinate with each other because they have no background knowledge to perform request allocation. Therefore, this case is suitable to apply the allocation cloud. The allocation cloud for video delivery services contains efficient request allocation algorithms and regard these algorithms as services. The allocation cloud collects the information of video requests and available video delivery services. Through actively detecting the video requests and the available video delivery services, the allocation cloud can decide feasible allocation services, such as COCA, and negotiate with consumers for request allocation. Since the allocation cloud can be aware of how consumer coordination impacts the profit of service providers, it can negotiate with service providers for consumers' addition benefits, such as video discount.

B. The Relation between Consumer Coordination and Service Providers' Profit

Experiment Setting: It is assumed that a video delivery service owns 500 videos and the average delivery time of a video is 30 minutes. The allocation cloud predicts the service requesting situation in the next one hour. The request arrival rate is from 5 to 30 requests per minute and is based on a Poisson distribution. The maximum waiting period of each request is between 2.5 minutes and 7.5 minutes and is based on an exponential distribution. It is assumed that there are four available allocation services,

Table I

COMPARISON THE DIFFERENCE OF RENEGE RATE FOR STRATEGIES WITHOUT COCA AND WITH COCA IN THE SIMULATION RESULT. RENEGE RATE IS THE NUMBER OF RENEGED REQUESTS DIVIDED BY THE TOTAL NUMBER OF REQUESTS MADE IN THE EXPERIMENT.

Channels	On-Demand			PRC		
	Without COCA	With COCA	Difference	Without COCA	With COCA	Difference
100	60%	34%	26%	56%	29%	27%
120	55%	34%	21%	51%	27%	24%
140	52%	30%	22%	46%	26%	20%
160	46%	29%	17%	42%	24%	18%
180	41%	26%	15%	37%	23%	14%

Table II

COMPARISON THE IMPROVEMENT OF SATISFIED REQUESTS FOR STRATEGIES WITHOUT COCA AND WITH COCA IN THE SIMULATION RESULT. SATISFIED REQUESTS IS THE NUMBER OF FULFILLED REQUESTS IN THE EXPERIMENT.

Channels	On-Demand			PRC		
	Without COCA	With COCA	Improvement (%)	Without COCA	With COCA	Improvement (%)
100	765	1227	60.39%	824	1326	27%
120	844	1231	45.85%	907	1354	24%
140	915	1301	42.50%	1004	1390	20%
160	1013	1330	31.29%	1330	1423	18%
180	1094	1375	25.69%	1375	1442	14%

On-Demand strategy [14], Pure-Rate-Control (PRC) strategy [15], On-Demand + COCA [2], and PRC + COCA [2]. The allocation cloud simulates the request allocation and the result is shown in Table I and Table II. Table I lists the renege rate in different channel numbers. The renege rate is the ratio of requests finally cancelled by service consumers due to over request deadline. Table II shows the number of satisfied requests in different channel numbers. According to the renege rate and satisfied requests in Table I and II, both on-demand allocation and PRC allocation can improve the number of satisfied requests by applying COCA. The more requests served, the higher profit gained by the service provider. Moreover, the profit is from consumers' coordination. Table I also displays the reduction of renege rate due to consumers' coordination. When video delivery service provides 100 channels, COCA reduces renege rate by 26% on the on-demand strategy and 27% on the PRC strategy. The reduction on the on-demand strategy and the PRC strategy is similar in different number of channels. Table I also shows that the reduction is more notable when the number of available channels is smaller. Table II displays the improvement of satisfied requests due to consumers' coordination. The result shows that satisfied requests can be improved approximately 60% when the video delivery service provides 100 channels. When the video delivery service provides 180 channels, the improvement is still over 20%.

Applying COCA to the Allocation Cloud: In brief, the result shows that video delivery services can have better satisfied requests but video requests with larger resource needs have to be reserved until the deadline. This indicates that some requests have to sacrifice their average waiting time to improve the total satisfied requests. Without the allocation cloud, service consumers cannot coordinate with each other and are not aware of their contributions on improving the profit of the video delivery service. With the allocation cloud, these service consumers who sacrifice their average waiting time can be aware of their contributions. The allocation cloud can also make favourable service agreements with video delivery service providers. Therefore, service consumers can reduce service price if they can afford to sacrifice their waiting time.

V. CONCLUSION AND FUTURE WORK

We propose the allocation cloud to deal with the request allocation for consumers. Consumer satisfaction can be improved through the allocation cloud because it optimizes service utilization and guarantees QoS. We introduce the operations of the allocation cloud and demonstrate its effectiveness through an experiment. According to the experiment results, we conclude that the allocation cloud collects service requests and reveals the relation between consumers' coordination and service execution. Therefore, service consumers can be aware of how to improve their satisfaction through feasible coordination. In the future, the consideration of consumer satisfaction optimization must be further explored. For this purpose, we need to decide a specific service and a set of consumers which make requests for the specific service. According to the specific service and corresponding consumers, allocation algorithms are selected in the allocation cloud. We plan to explore the consumer satisfaction optimization based on these allocation algorithms.

REFERENCES

[1] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "Qos-aware middleware for web services composition," *IEEE Trans. Softw. Eng.*, vol. 30, no. 5, pp. 311–327, May 2004.

[2] Y.-W. Ho, "Coordinated resource allocation for video delivery services," Master's thesis, National Cheng Kung University, Tainan, Taiwan, 2009.

[3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Comm. ACM*, vol. 53, no. 4, pp. 50–58, Apr. 2010.

[4] M. Creeger, "CTO Roundtable: Cloud computing," *Comm. ACM*, vol. 52, no. 8, pp. 50–56, Aug. 2009.

[5] A. Ojala and P. Tyrväinen, "Developing cloud business models: A case study on cloud gaming," *IEEE Softw.*, vol. 28, no. 4, pp. 42–47, Jul./Aug. 2011.

[6] P. Louridas, "Up in the air: Moving your applications to the cloud," *IEEE Softw.*, vol. 27, no. 4, pp. 6–11, Jul./Aug. 2010.

[7] M. Cusumano, "Cloud computing and SaaS as new computing platforms," *Comm. ACM*, vol. 53, no. 4, pp. 27–29, Apr. 2010.

[8] F. I. Popovici and J. Wilkes, "Profitable services in an uncertain world," in *Proc. 2005 ACM/IEEE Conf. Supercomputing (SC'05)*, Seattle, USA, Nov. 2005.

[9] K. Tsakalozos, H. Kllapi, E. Sitaridi, M. Roussopoulos, D. Paparas, and A. Delis, "Flexible use of cloud resources through profit maximization and price discrimination," in *Proc. 2011 IEEE 27th Int'l Conf. Data Eng. (ICDE'11)*, Hannover, Germany, Apr. 2011, pp. 75–86.

[10] V. Kantere, D. Dash, G. François, S. Kyriakopoulou, and A. Ailamaki, "Optimal service pricing for a cloud cache," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1345–1358, Sep. 2011.

[11] J. Chen, C. Wang, B. B. Zhou, L. Sun, Y. C. Lee, and A. Y. Zomaya, "Tradeoffs between profit and customer satisfaction for service provisioning in the cloud," in *Proc. 20th Int'l Symposium on High Performance Distributed Computing (HPDC'11)*, San Jose, USA, Jun. 2011, pp. 229–238.

[12] A. Dan, D. Sitaram, and P. Shahabuddin, "Scheduling policies for an on-demand video server with batching," in *Proc. Second ACM Int'l Conf. Multimedia*, San Francisco, California, USA, Oct. 1994, pp. 15–23.

[13] A. Dan, D. Sitaram, and P. Shahabuddin, "Dynamic batching policies for an on-demand video server," *Multimedia Systems*, vol. 4, no. 3, pp. 112–121, Jun. 1994.

[14] Y. Zhang, M.-Y. Wu, and W. Shu, "Adaptive channel allocation for large-scale streaming content delivery systems," *Multimedia Tools and Applications*, vol. 32, no. 3, pp. 253–273, Mar. 2007.

[15] K. C. Almeroth, "Adaptive workload-dependent scheduling for large-scale content delivery systems," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 426–439, Mar. 2001.