Incorporating Human Psychological Factor in Assessing the Deliverability of Quality of Service (QoS) for Multimedia Content Adaptation Services

Mohamed J.[1,2], Fudzee M. F. M.[2] and Ismail M.N.[2]

[1]*Kolej Poly-Tech MARA Batu Pahat, Johor*
[2]*Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn, Malaysia*

ABSTRACT

Content adaptation is a potential solution for tailoring multimedia web content according to the users' preferences and heterogeneous devices' constraints. Content adaptation can be done as third party service over the Internet. Users may pay for the service thus demand quality. The quality should include the human psychological factors. One of these factors is the maximum time a user can wait for the output to be displayed. Thus, response time is one of the qualities of service (QoS) to be considered in assessing the deliverability of content adaptation services. However, the advertised response time may not be deliverable accordingly during the actual service execution due to heavy load. Practically, the service provider should able to determine a current deliverable response time before the service level agreement (SLA) is settled with the users. In this paper, we propose a strategy for service providers to evaluate incoming requests and capable of offering the new response time. The proposed strategy takes into account the current server load and enables a mechanism for the user to evaluate whether the new response time can be accepted or not. We analyzed the performance of the proposed strategy in terms of SLA settlement under various conditions. The results indicate that the proposed strategy performs well.

Keywords: multimedia content adaptation, QoS, SLA, Service-oriented content adaptation.

I INTRODUCTION

Content adaptation has in fact already gained considerable importance in today's multimedia communications and will certainly become an essential functionality of any service, application or system in the near future (Carreras *et. al*, 2009). As the heterogeneity of mobile device such as smart phone and tabs growth, searching websites through mobile device is also demanding. The recent trend is that content adaptations are largely demanded and became profitable mobile contents of its quality to the Internet as third party services (Arai & Tolle, 2010). Existing content adaptation systems employing these approaches such as client-side, server-side or proxy-side adaptation (Malandrino *et. al.*, 2010) operate in isolation, often encounter limited adaptation functionality, get overload if too many concurrent users and open to single point of failure, thus limiting the scope and scale of their services (Yang & Shao, 2007).

Digital multimedia services and devices are now mediated through increasingly proportion of human activities such as social interactions, entertainment, shopping, and gathering information (Kosinski, Stillwell & Graepel, 2013). To get more accuracy and rapid information, users may pay for the service thus demand quality. Generally, the response of those services is not very immediately but rather delayed to a certain extent depends on the type of request and the type of desired response. At the end, these response times is largely dominated to user-perceived quality (Egger *et. al*, 2012).

The idea of establishing content adaptation as a service to allow potential solution for tailoring multimedia web content according to the users' preferences and heterogeneous devices' constraints has been endorsed. Typically, there are many service providers offering a variety of content adaptation that can be suit. Therefore, the solution for these services is to cooperate with each other through service composition to completely serve the request that they cannot attain individually. This is termed as Service-oriented Content Adaptation (SOCA). SOCA presents a model to enable content adaptation to be consumed as Web services (Fudzee & Abawajy, 2011). However, current solutions from (Alhamad, Dillon & Chang, 2010) do not take into account the human psychology in negotiating with services.

Figure 1: SOCA Framework shows the architecture of service level agreement (SLA) by SOCA framework. This system is based from the service oriented architecture (SOA). The purpose of the SOCA framework is to give out user expectation by offering added value of content and also to provide flexible and scalable service-based content delivery mechanism. The framework consists of components that provide access to content servers, develop user request to source format, manage and provide

content description (Fudzee, Abawajy & Deris, 2010).

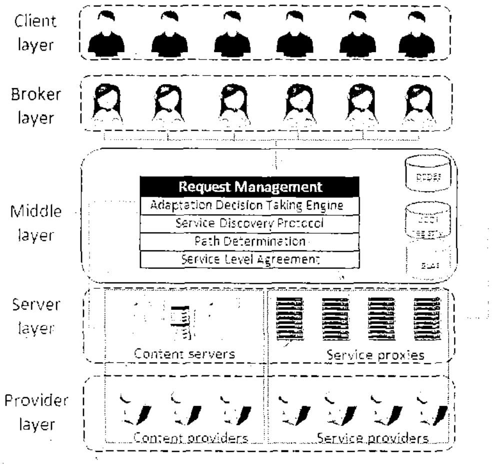

Figure 1: SOCA Framework

Commonly, SOCA enables an adapted content to be delivery by a number of services; service composition based on quality of service (QoS) is known to be a practical solution. Hence, response time is one of the qualities of service (QoS) to be considered in assessing the deliverability of content adaptation services. However, the advertised response time may not be deliverable accordingly during the actual service execution due to heavy load.

The quality of mobile services in content adaptation should include the human psychological factors. One of these factors is the maximum time a user can wait for the output to be displayed. Lobo *et. al*, (2011) stress that there will be more frustration while making longer enough of time to wait for pages to load, so as a result they usually quit the task or try to find information to another site. This is also relevant to content adaptation service response time.

Practically, the service provider should able to determine a current deliverable response time before the service level agreement (SLA) is settled with the users. This strategy has been found out from (Md Fudzee & Abawajy, 2011) by proposed QoS negotiation modeling of response time by considering one to one negotiation strategy (from broker as user on behalf to service provider).

In this paper, we propose a strategy for service providers to evaluate incoming requests and capable of offering the new response time. The proposed strategy takes into account the current server load and enables a mechanism for the user to evaluate whether the new response time can be accepted or not. We analyzed the performance of the proposed strategy in terms of SLA settlement under various conditions. The results indicate that the proposed strategy performs well. This paper consist of six section which includes our related work in section 2, the proposed solution in section 3, the performance evaluation in section 4, result and discussion in section 5 and the conclusion for this paper in section 6.

II RELATED WORK

Previous studies have overcome the problems of service provider from SOCA framework. SOCA comes with existing one to one negotiation strategy (Md Fudzee & Abawajy, 2011) that is most works dealing with client-side quality evaluation perform recommendations during service selection process. In this strategy, the waiting time has been focusing to reflect the response time. The broker reaches the requested service by searching the priority function. The current server load checks by SLA assessor and estimates the value of request it can be served within offered waiting time.

This strategy offered single session task by SLA assessor to be consider by waiting time and at the same time the new waiting time for the request that might be rejected. Figure 2: One-To-One Negotiation depicted an example of implementing *E(W)* QoS negotiation between brokers with a content adaptation provider.

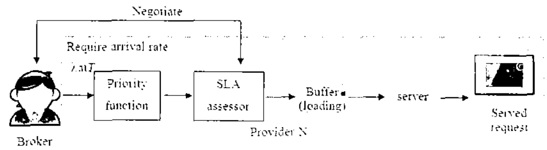

Figure 2: One-To-One Negotiation

From the Figure 2: One-To-One Negotiation, once the agreements are entered into, it becomes necessary to monitor their conditions, which will commonly relate to timeliness, reliability and request throughput, at run time (Raimondi, Skene & Emmerich, 2008). The problem with this strategy is broker should have to stick around for a few seconds to receive the adapted content. As a result, there will be a heavy load of incoming request, false of information and disclaimed information. To overcome this problem, we propose a strategy for service providers to evaluate incoming requests and capable of offering the new response time. The proposed strategy called One-to-Multiple negotiation. This strategy takes into account the current server load and enables a mechanism for the user to evaluate whether the new response time can be accepted or not.

III PROPOSED SOLUTION

This section, we will describe about the proposed of One-to-Multiple Negotiation strategy that has capability to determine that SLA can be settled with overall waiting time for all task rather than a single task. This improvement strategy based from (Al-Jaljouli & Abawajy, 2012) will consider less of response time by applying overall task to a certain request. Figure 3: One-To-Multiple Negotiation Strategy illustrates the one-to-multiple negotiation strategy.

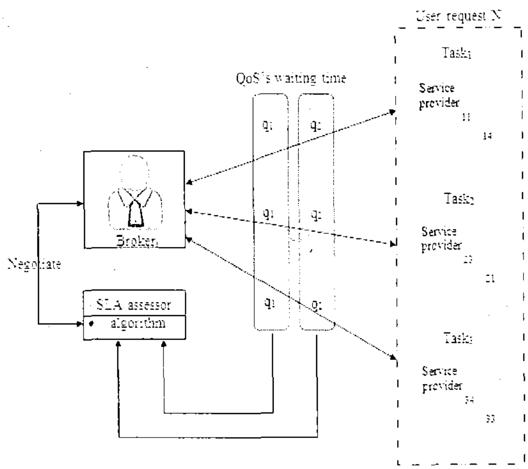

Figure 3: One-To-Multiple Negotiation Strategy

Assuming the $broker_i$ request several tasks namely $task_1$, $task_2$ and $task_3$. Next, the SLA waiting time (QoS) will be considered as 3 seconds. The service provider will check the current load and calculating the request that can be served. Further assume the service provider will be served by two options. The first option in $task_1$, $task_2$ and $task_3$ offered exactly 3 seconds and the second option will offer 3 seconds in $task_1$, 2 seconds in $task_2$ and 4 seconds in $task_3$. The estimation of QoS waiting time describes thru Table 1: Qos Waiting Time Estimations.

Table 1: Qos Waiting Time Estimations

	$Task_1$	$Task_2$	$Task_3$
1st option	3 sec.	3 sec.	3 sec.
2nd option	3 sec.	2 sec.	4 sec.
Consideration	Accept	Accept	Re-offer

These tasks will accept all the requested needs from the service provider by replying back any feedback within the two options given. All tasks will take maximum of 3 seconds per services to fulfill the settlement of SLA. Typically, all the tasks completed in terms of technically. Therefore, this strategy can allow user to get any responds from several task and it will decrease user's perception about waiting time. By applying this method, the enhance user experience of the content adapted and also provide flexible and scalable service-based content delivery mechanism. An algorithm is applied to this strategy at SLA assessor for improvement settlement rate of waiting time factor considers by QoS. Table 2: Common Symbol list the common symbol used for this paper based on proposed model.

Table 2: Common Symbol

Symbol	Description
N	Arrival rate of request
$E(W)$	Expected waiting time
S	Server load

SLA Assessor

The inputs for the enhancement of SLA assessor algorithm pioneered from Liu *et. al,* (2010) are the requests inclusive the current server load, the QoS and their priorities. Figure 4: Algorithm Of SLA Assessor states the algorithm by communication steps of settlement SLA by SLA assessor.

```
Algorithm 1: SLA Assessor
INPUT: server load
OUTPUT: settlement of SLA
BEGIN
1: Retrieve requests and their arrival with server load.
2: estimation of total requests can be served within
   the advertised E(W)
3: FOR each request DO
4:   IF request can be served within advertised
     E(W)
     THEN
5:     Settle agreement
6:   ELSE
7:     Check uncompleted tasks and schedule it first.
7.1        Calculate the sub-deadlines for tasks of
           the last instance.
7.2        Calculate the sub-deadlines for tasks of
           other instances based on the last
           instance assuming that the schedule
           follows the stream-pipe mode to avoid
           competitions for lower services.
8:     Calculate the estimated execution time for each
       service
9:     Allocate each task to the service which gives the
       execution time that does not exceed the sub-
       deadline.
10:    Provide a just-in-time time-cost relationship
       graph for user to optionally choose an updated
       compromised deadline for scheduling.
11:    Settle agreement if accepted by the broker
12:  END IF
13: END FOR
END
```

Figure 4: Algorithm Of SLA Assessor

Specifically, during the QoS negotiation process, the waiting time QoS E(W) of each content adaptation server will be evaluated by manipulating Pollaczek-Khintchine formula as in (Pathan & Buyya, 2009):

$$E(W) = \lambda E(X^2)/2(1-s) \qquad \text{Eq. (1)}$$

Next, $E(X)$ clarifies as the mean service time. A Bounded Pareto adopted from (Pathan & Buyya, 2009) allocation followed of each incoming request size. A Bounded Pareto B (k,ρ,α) of probability density function (Pathan & Buyya, 2009) is:-

$$f(x) = \frac{\alpha k^{\alpha}}{1-(\frac{k}{s})^{\alpha}} x^{-\alpha-1} \qquad \text{Eq. (2)}$$

where the task size variation represents by α, k and s which are smaller and the largest possible task size with $(k \leq x \leq s)$. High variability is observed when $(\alpha \approx 1)$ and when $(\alpha \approx 2)$ observed by moderate variability. The expected waiting time of $E(W)$ is predict before a request being serviced so that the requests potentially being rejected by offering new service time of x. $E(X)$ of the mean service time issues by the following mathematical formula (Pathan & Buyya, 2009):

$$E(X^j) = \begin{cases} \frac{\alpha k^{j} (\frac{k}{s})^{\alpha} - (\frac{k}{s})^{j}}{(j-\alpha)(1-(\frac{k}{s})^{\alpha})} & if\ j = \alpha \\ \frac{\alpha k^{\alpha} \ln(\frac{s}{k})}{(1-(\frac{k}{s})^{\alpha})} & if\ j = \alpha \end{cases} \qquad \text{Eq. (3)}$$

The current load estimations within given number of the requests by plotting $E(W)$. For instance, the waiting time of QoS had been advertised by three service provider namely *task_1*, *task_2* and *task_3* by assuming the fair load of server (0.6) and the N as 50. The $E(W)$ estimated by SLA assessor and only form 30 to 80 requests can be realized to be served by $E(W)$ being offered. As a result, the SLA is settled with the first 30 until 70 of requests by *task_1* of the overall task.

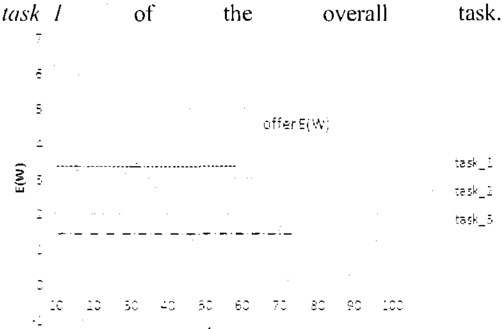

Figure 5: offered *E(W)* against three tasks based on current load.

From the : offered ***E(W)* against three tasks based on current load.**, assuming that the *E(W)* evaluates by SLA assessor for the rest of requests. Further assume that the first of 30 requests is offered 3.3 x 10^9 for the *E(W)*, then the rest of 70 requests offered with 1.5 x 10^9. The SLAs are settled if the broker accepts the *E(W)* by negotiating at *task_1* otherwise the *task_2* and *task_3* will continuing the journey to be received the requests. From this, we can conclude that the overall task can be agreed by both of the broker and service provider referring to the service level agreement.

IV EVALUATION OF PERFORMANCE

We use the simulation to study the efficiency of the service provider selection by focusing on the expected waiting time of *E(W)* by takes into account the current server load and enables a mechanism for the user to evaluate whether the new response time can be accepted or not.

To improve the performance of expectation waiting time, the current methodology from (Fudzee & Abawajy, 2011) has been followed. Table 3: Simulation settings shows the simulation setting and the parameter to this study.

Table 3: Simulation settings

Parameter	Value
λ	50 to 100
α	2
ρ	0.5 (lightly load) to 0.9 (heavy load)
k	1010.15
s	10^{10}
service distribution	$f(x) = \frac{\alpha k^{\alpha}}{1-(\frac{k}{s})^{\alpha}} x^{-\alpha-1}$, $(k \leq x \leq s)$
E(W) acceptance rate	0% to 30%

The improvement approach based from (Zheng, Zhang, & Lyu, 2010) and (Zhang, Zheng, & Lyu, 2011) gives more implication to the SLA. We analyzed the performance of the proposed strategy in terms of SLA settlement under various conditions. The results indicate that the proposed strategy performs well.

V RESULT AND DISCUSSION

Referring to the Figure 6: SLA acceptance rate versus requested numbers (default server load 0.6), the SLA acceptance rate shows at y-axis and the function of the number requested show by x-axis. We transforms N of incoming request from 50 to 100 from the simulation settings. The server load in this case takes into account as 60 by a fair load of 0.6 advertised by *E(W)*. The acceptance rate of QoS offered by *E(W)* spread of 0% to 30 %.

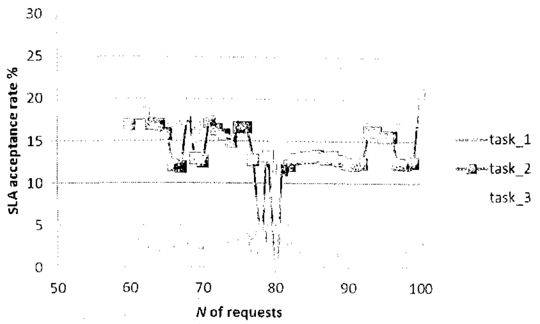

Figure 6: SLA acceptance rate versus requested numbers (default server load 0.6)

As in Figure 6: SLA acceptance rate versus requested numbers (default server load 0.6), there are three task calls for a request, *task_1, task_2* and *task_3*. From the server load at 60, the three tasks play role to give their service with difference number. There are offered the dramatically lowest request from *task_1* while the other two tasks *(task_2 and task_3)* offer considerably higher chances to run the request along at x-axis. The requested along x-axis increases about 23% at *task_3* then at *task_1*. This simulation proves that *task_3* can served with higher cumulative rate of *E(W)*. As a result, the settlement rate will be more flexible due to the higher cumulative of *E(W)*. The expectation for the simulation is when the settlement rate is higher so that the SLA is settled refer to the agreement. This is because the strategy can reduce the upcoming request from second option for each task once the first option cannot satisfy with advertised waiting time.

VI CONCLUSION

Content adaptation is a potential solution for tailoring multimedia web content according to the users' preferences and heterogeneous devices' constraints. Content adaptation can be done as third party service over the Internet. User consumed that the adaptation service can pay them good service. The quality should include the human psychological factors. One of these factors is the maximum time a user can wait for the output to be displayed. Thus, response time is one of the qualities of service (QoS) to be considered in assessing the deliverability of content adaptation services. Practically, the service provider should able to determine a current deliverable response time before the service level agreement (SLA) is settled with the users. We summarized our contribution into: (1) we proposed the improvement of SLA assessor tailor to the SOCA framework, (2) the proposed strategy for SOCA framework simulated by several tasks. In future, we plan to study on monitoring of the strategy by serve a given task advertised of QoS.

ACKNOWLEDGMENT

The authors would like to acknowledge Universiti Tun Hussein Onn Malaysia and Ministry of Education, Malaysia for the Multi-Disciplinary Research Grant 1107 provided in financing the research activities.

REFERENCES

Al-Jaljouli, R., & Abawajy, J. (2012). Strategies for Agent-Based Negotiation in E-Trade. Network and traffic engineering in emerging distributed computing applications, 43-65.

Alhamad, M., Dillon, T., & Chang, E., 2010. Sla-based trust model for cloud computing. In Network-Based Information Systems (NBiS), 2010 13th International Conference on (pp. 321-324). IEEE.

Arai, K., & Tolle, H. (2010). Automatic e-comic content adaptation. International Journal of Ubiquitous Computing, 1(1), 1-11.

Carreras, A., et. al., 2009. Context-aware and DRM-enabled content adaptation platform for collaboration applications. IEEE Multimedia.

Egger, S., Hossfeld, T., Schatz, R., & Fiedler, M. (2012, July). Waiting times in quality of experience for web based services. In Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on (pp. 86-96). IEEE.

Fudzee, M. F. M., & Abawajy, J. H. (2011). A Strategy of Negotiating QoS for Content Adaptation Services. Deakin University, School of Information Technology.

Fudzee, M. F. M., Abawajy, J., & Deris, M. M. (2010, May). Multi-criteria content adaptation service selection broker. In Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on (pp. 721-726). IEEE.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802-5805.

Liu, K., Jin, H., Chen, J., Liu, X., Yuan, D., & Yang, Y. (2010). A compromised-time-cost scheduling algorithm in swindew-c for instance-intensive cost-constrained workflows on a cloud computing platform. International Journal of High Performance Computing Applications, 24(4), 445-456.

Lobo, D., Kaskaloglu, K., Kim, C. Y., & Herbert, S. (2011). Web usability guidelines for smartphones: A synergic approach. International journal of information and electronics engineering, 1(1), 33-37.

Malandrino, D., Mazzoni, F., Riboni, D., Bettini, C., Colajanni, M., & Scarano, V. (2010). MIMOSA: context-aware adaptation for ubiquitous web access. Personal and Ubiquitous Computing, 14(4), 301-320.

Md Fudzee, M. F., & Abawajy, J. H. (2011). QoS-based adaptation service selection broker. Future Generation Computer Systems, 27(3), 256-264.

Pathan, M., & Buyya, R. (2009). Architecture and performance models for QoS-driven effective peering of content delivery networks. Multiagent and Grid Systems, 5(2), 165-195.

Raimondi, F., Skene, J., & Emmerich, W., 2008. Efficient online monitoring of web-service SLAs. In Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering (pp. 170-180). ACM.

Yang, S. J., & Shao, N. W. (2007). Enhancing pervasive Web accessibility with rule-based adaptation strategy. Expert Systems with Applications, 32(4), 1154-1167.

Zhang, Y., Zheng, Z., & Lyu, M. R. (2011, October). Exploring latent features for memory-based QoS prediction in cloud computing. In Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on (pp. 1-10). IEEE.

Zheng, Z., Zhang, Y., & Lyu, M. R. (2010, July). Distributed qos evaluation for real-world web services. In Web Services (ICWS), 2010 IEEE International Conference on (pp. 83-90). IEEE.