Associating User’s Psychology into Quality of Service: An Example of Web Adaptation Services

Mohd Farhan Md Fudzee, Juliana Mohamed, Azizul Azhar Ramli, Shahreen Kasim, Hairuinizam Mahdin, Mohd Norasri Ismai, Azizan Ismail & Mohd Hanif Jofri

Faculty of Computer Science and Information Technology University Tun Hussein Onn Malaysia Parit Raja, 86400 Batu Pahat, Johor, Malaysia (farhan,azizulr, shahreen, hairuln, norasri, mhanif)@uthm.edu.my

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article history:
Received 22 January 2017
Accepted 03 February 2017
Available online 05 February 2017

Index Terms:
Quality of service, Web adaptation service, User psychology

ABSTRACT

Content adaptation has 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 to take 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.

1. 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 (Anna 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 (Ami & Tolle, 2010). Existing content adaptation systems employing these approaches such as client-side, server-side or proxy-side adaptation (Malandrinou 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 these 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, Hossfeld, Schatz, & Fiedler, 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 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)

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, Kaskaloglu, Kim & Herbert (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 (Fudzee, 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 section 5 with conclusion section 6.
section 4, result and discussion in section 5 and the conclusion for this paper in section 6.

2. Literature Review

Previous studies have overcome the problems of service provider from SOCA framework. SOCA comes with existing one to one negotiation strategy (Fudzee, 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 depicted an example of implementing E(W) QoS negotiation between brokers with a content adaptation provider.

From the Figure 2, 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 & Emmrich, 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.

3. General MCDM Design Model and Methodology for Teaching Performance Evaluation

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. This improvement strategy (Fudzee, 2011) that is most works dealing with client-side quality evaluation perform recommendations during service selection process.

The broker reaches the requested service by searching the priority function. The current load estimations within given number of the requests by SLA assessor and estimates the value of request it can be served within offered waiting time. As a result, the SLA is settled with the first 30 until 70 of requests can be realized to be served by E(W) being plotted 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 largest possible task size with (k≤ x ≤ s). High variability is observed when (α approx 1) and when (α 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):

\[ E(X) = \frac{1}{\lambda(1+\alpha)} \]

where the task size variation represents by α, k and s which are smaller and the largest possible task size with (ks x s). High variability is observed when (α approx 1) and when (α approx 2) observed by moderate variability. The expected waiting time of E(W) is predict before a request being served 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):

\[ E(X) = \frac{1}{\lambda(1+\alpha)} \]

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.

| Task | Task | Task |
|------|------|------|
| 1st option | 3 sec. | 2 sec. | 4 sec. |
| 2nd option | 3 sec. | 3 sec. | 2 sec. |
| Consideration | Accept | Accept | Reoffer |

Table 1: QoS waiting time estimations

Table 2: common symbol

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

Figure 2: One-to-one negotiation

SLA Assessor

The inputs for the enhancement of SLA algorithm pioneered from Liu et. al, (2016) are the requests inclusive the current server load, the QoS and their priorities. Figure 4 states the algorithm by communication steps of settlement SLA by SLA assessor.

Algorithm 1: SLA Assessor

1. INPUT: settlement of SLA
2. SLA Assessor
3. Receive request and their arrival with server load.
4. Estimate the total requests can be served within the advertised E(W)
5. FOR each request DO
   1. IF request can be served within advertised E(W) THEN
      1. Settle agreement
   2. ELSE SLA Assessor
   3. Check unprocessed tasks and schedule it first.
   4. Calculate the sub-admission for tasks that last but not the last one.
   5. Allocate each task to the service which gives the execution time that does not exceed the sub-admission
   6. Provide a job-scheduling-time relation graph for user to optionally choose an updated compromise function for time slotting
   7. Settle agreement if accepted by the broker
   8. END IF
   9. END
10. END FOR
11. FDB

Figure 4: algorithm of SLA assessor

Specifically, during the QoS negotiation process, the waiting time QoS E(X) of each content adaptation server will be estimated by manipulating the Erlang-A Kuma formula as in (Pathan, & Buyya, 2009):

\[ E(X) = \frac{1}{\lambda(1+\alpha)} \]

\[ F(X) = \frac{1}{\lambda(1+\alpha)} \]

\[ E(X)|t = \frac{n}{\lambda(1+\alpha)} \]

where t is the task size variation represents by α, k and s which are smaller and the largest possible task size with (ks x s). High variability is observed when (α approx 1) and when (α approx 2) observed by moderate variability. The expected waiting time of E[W] is predict before a request being served 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):

\[ E(X)|t = \frac{n}{\lambda(1+\alpha)} \]

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.
From the Figure 5, 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 109 for the $E(W)$, then the rest of 70 requests offered with 1.5 x 109. 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.

4. Conclusion

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 (Md Fudzee, & Abawajy, 2011) has been followed. Table 3 shows the simulation setting and the parameter to this study.

The improvement approach based from (Zhang, Zheng, & 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.

Referring to the Figure 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 %.

As in Figure 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 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 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 task_3.

### Table 3: Simulation settings

| Parameter | Value |
|-----------|-------|
| $k$       | 10^{10} |
| $\rho$    | 0.5 (lightly loaded) to 0.9 (heavy load) |
| $\sigma$  | 2 |
| $\alpha$  | $1 - \frac{
| $\beta$   | $\frac{1}{k^{1}}$ |
| $\gamma$  | $1 - \frac{1}{s}$ |
| $E(W)$ acceptance rate | 0% to 30% |

Figure 7: SLA acceptance rate versus requested numbers (default server load 0.5)

In future, we plan to study on monitoring of the strategy by serve a given task advertised of QoS. 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.

4. 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 context 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.

### References

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

Arai, K., & Tolle, H. (2010). Automatic e-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., 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., Huang, 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.

M.F. Md Fudzee & J. Abawajy. 2011. QoS-based content adaptation service selection broker. Elsevier's Future Generation Computer System 27 (3), 256-264.

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

Pathan, M., & Buyya, R. (2009). Architecture and performance models for QoS-driven effective peering of content delivery networks. Multisagent 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 SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.

S.H. Yang & N.W.Y Shao. 2007. Enhancing pervasive Web accessibility with QoS-based content adaptation service selection broker. In Proceedings of the 16th ACM SIGSOSP International Symposium on Foundations of software engineering (pp. 170-180). ACM.