Technical considerations towards mobile user QoE enhancement via Cloud interaction

Elias Allayiotis
School of Computing
UCLan
Preston, UK
EAllayiotis@uclan.ac.uk

Josephina Antoniou
School of Sciences
UCLan Cyprus
Pyla, Cyprus
JAntoniou@uclancyprus.ac.cy

***Abstract*— This paper discusses technical considerations of a Cloud infrastructure which interacts with mobile devices in order to migrate part of the computational overhead from the mobile device to the Cloud. The aim of the interaction between the mobile device and the Cloud is the enhancement of parameters that affect the Quality of Experience (QoE) of the mobile end-user through the offloading of computational aspects of demanding applications. This paper shows that mobile user's QoE can be potentially enhanced by offloading computational tasks to the Cloud which incorporates a predictive context-aware mechanism to schedule delivery of content to the mobile end-user using a low-cost interaction model between the Cloud and the mobile user. With respect to the proposed enhancements, both the technical considerations of the cloud infrastructure are examined, as well as the interaction between the mobile device and the Cloud.**

Keywords*— *Cloud Architecture; Mobile to Cloud interaction; Mobile QoE; Mobile Cloud Computing

I. INTRODUCTION & RELATED WORK

A. Current Mobile & Cloud Computing Environment

Nowadays, smart mobile devices are becoming the computing devices of choice for delivering information to people around the world either for business use, entertainment or personal use. With the convergence of the mobile phone and the personal computer within the same device, the demand for fast and high-performing mobile devices is increasing rapidly.

Cloud computing (Cloud) offers computation as a service. It enables a flexible, scalable and cost-effective computational model where complicated infrastructure setups are abstracted from the Cloud user and provided as a service. Cloud computing provides large amounts of storage space and delivers demanding computational tasks in a fast and affordable way. This is achieved by distributing computational loads within a network of interconnected devices so that sets of computational operations are executed in parallel.

Since the first Public Cloud offering in 2006 by Amazon Web Services, there has been an aggressive push by large technology organizations to build data-centers offering Cloud services across the world [1]. Public Cloud is considered an enabling tool for affordable storage and processing power for anyone with an Internet connection. Its pay-as-you-go service model provides computing professionals with a flexible, scalable and on-demand computing environment eliminating the risk of upfront infrastructure expenditures and lengthy infrastructure setups.

B. Cloud Provider vs. Cloud User

From the perspective of the Cloud Provider, Cloud computing becomes a multi-tier optimization problem with the end-goal of making data-centres efficient in both hardware and software. In terms of hardware, Cloud Providers focus in eliminating cost from their operations by minimizing operational overheads (electricity, machine-cooling costs) [2]. In terms of software, Cloud Providers focus in improving data processing speeds and eliminate waste from their network communications by re-designing and improving data interchange. This paper focuses on the software aspect by examining functional architecture considerations for enhancing data processing which can potentially lower costs for the Cloud Provider.

From the perspective of Cloud Users, and in particular mobile device users, the central issue becomes their experience with the service (QoE). As defined in existing literature, QoE can be a subjective measurement of a user's experience with a service [3] as well as an objective measurement of the environment's quality of service (QoS) parameters [4]. Furthermore, is has been shown that the user's context (location. mobility, etc.) can be used not only to define QoE but also predict it [5]. This paper does not offer a precise definition of a mobile user's QoE rather than attempts to address two factors that have been directly linked with QoE performance. The first factor is the speed of information delivery to the mobile user and the second factor is the relevance of the information delivery to the user based on the user's current context.

In future work, we aim in defining QoE for mobile device users as a combined function of all of these different parameter types (Section IV, B). However, for the purpose of this paper, QoE refers only to the speed of information delivery and the relevance of that information to the user, as stated above.

Considering the limited processing power of mobile devices and the tremendous processing capability of the Cloud, this paper explores the scenario of offloading computational tasks from a mobile device to the Cloud using a low-cost communication model. Furthermore, it shows that an implementation of this scenario can potentially enhance the mobile user's QoE and also provide an efficient context-aware Cloud processing mechanism.

C. *Mobile Computing Considerations*

There is a vast amount of work that has been done in the areas of Mobile Networking, Context-Awareness and QoE of mobile devices. These areas are now being adopted and re-worked using the Cloud as an offloading mechanism and an optimization vehicle. These approaches are starting to shape the field of Mobile Cloud Computing. As this paper explores a Mobile-to-Cloud interaction, some further considerations and research opportunities from Context-Aware Mobile Computing are identified.

1) Multiparty Networking: Research that has been conducted in Multiparty Networking [6] and in architectures for Context Multicast delivery [7] aims in improving user experience, as can be quantified through QoE measures, by allowing mobile networks to support context-aware information. This model provides a heterogeneous networking environment where the Cloud can be considered as an additional optimization element.

2) Session Management: Session Management and in specific service migration of ad-hoc networks [8] aim in optimizing interactions between the mobile client and a service using context-sensitive binding. Session management needs to be carefully considered as a way to integrate potential enhancements in context-aware environments. Furthermore, Cloud infrastructure can be considered as a potential candidate for optimizing process migration in such an environment.

II. ARCHTECTURE CONSIDERATIONS

A. *Orchestration of Resources*

The proposed setup considers enhancing user QoE through a functional separation of the computational aspects of content delivery, from the aspects of content delivery directly experienced by the user, such as interface aspects with the processed content. Refining content delivery is achieved by the aid of the Cloud, through offloading procedures. Furthermore, both the Cloud side and the mobile device side of the offloading process can be further enhanced by considering such aspects as context-awareness of user parameters as well as infrastructure-specific context, e.g. availability and bandwidth.

Regarding context, the proposed scheme aims in achieving enhanced context-awareness by utilizing this functional separation to better support context-aware information, where context of the user, session, network and environment is carefully considered. The proposed technical considerations and functionalities are described through a specific application to a context-driven use case scenario (Section II, B). We show that the proposed setup is potentially able to provide personalized QoE enhancements to the users, in a context-aware manner. Next, we discuss the proposed setup in terms of resource orchestration in more detail.

The proposed Cloud setup is divided into four linked layers:

- Data Receiving layer,
- Data Processing layer,
- Data Storing layer, and
- Data Sending layer.

An overview of the orchestration of resources is illustrated in Fig. 1.

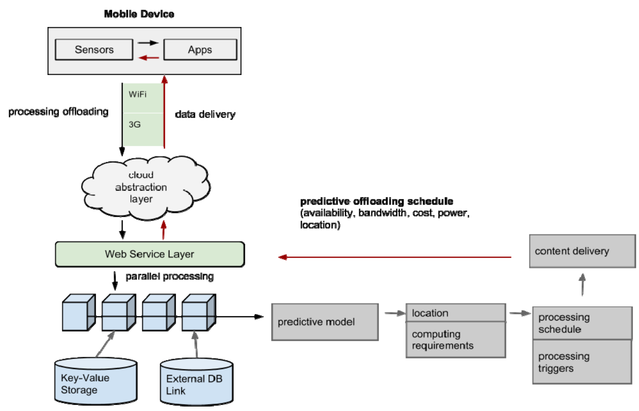

Fig. 1. Orchetration of resources

The following subsections provide further detail on each layer.

1) Data Receiving Layer: The data receiving layer is composed of a RESTful web service [9] which accepts GET and POST requests from the mobile device. All requests are received from the data receiving layer and are sent to the Data processing layer.

Depending on the Public Cloud provider which is chosen for implementation, certain load balancing mechanisms can be deployed between the data receiving and the data processing layer in order to ensure high availability of the service. As illustrated in Fig. 2, the Elastic Load Balancer [10] of Amazon Web Services is used to automatically scale requests to multiple cloud instances.

2) Data Processing Layer: The data processing layer utilizes a predictive model which takes into account the context attributes plus the requested data and develops three things:

- The data processing schedule,
- The data processing triggers and

- The requested content to be delivered.

The Data Processing layer holds a series of Data Processing Nodes. The Data Processing Node, as illustrated in Fig. 3, is designed in such a way that it can be executed in parallel and utilize the full capabilities of the Cloud infrastructure.

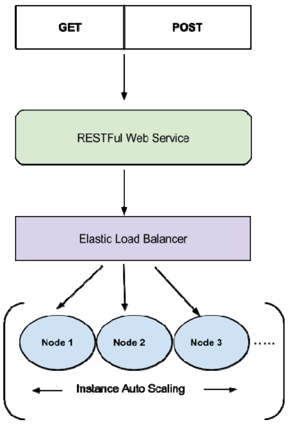

Fig. 2. Data Receiving Layer

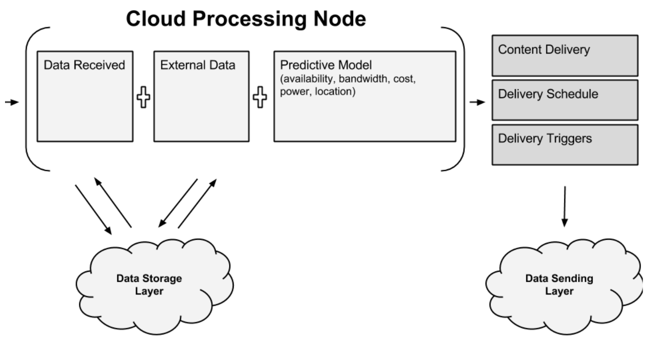

Fig. 3. Data Processing Layer

3) Data Storing Layer: The results from processing (Data Processing Layer) are then sent to the Data Storing layer to be archived in the user's database. The structure of the data is in key-value pairs and utilizes a MapReduce mechanism which is a very efficient and "embarrassingly parallel" operation [11]. The Data Storing Layer holds a series of Data Storing Nodes which run an implementation of MapReduce. A high level illustration of the Data Storing node is shown in Fig. 4.

4) Data Sending Layer: Finally, the results are being sent to the Data sending layer, as illustrated in Fig. 5, which compresses the data and structures them in JSON [12]. It has been shown in existing literature that compressed JSON format data via a RESTFul interface is an efficient way of transporting data from the cloud to a mobile device [13].

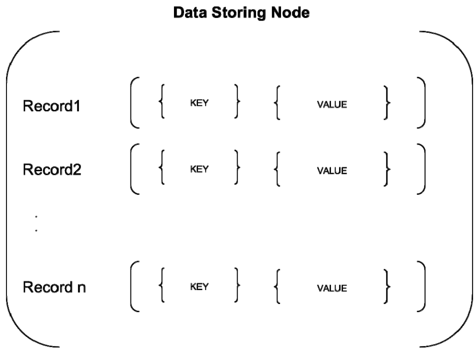

Fig. 4. Data Storing Layer

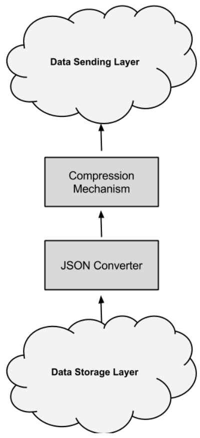

Fig. 5. Data Sending Layer

B. *Use Case Scenario*

To better illustrate the proposed technical enhancements and communication interactions, the following example is considered as a potential use case scenario.

A mobile user is a business person who travels out of town often for business meetings. Before every meeting, the user searches online via a smartphone to find a local copy center with Internet-enabled workstations to download and print various materials.

Since meetings are usually out of town, the user connects to the Internet either through an open WiFi hotspot (if available) or a 3G connection. There are times that the network is busy and results are too slow to load on the device. In addition, the user gets charged out of town data roaming fees when using a 3G connection.

Consequently, the QoE of the end user is compromised by a slow Internet connection while the user's cost is increased due to data roaming charges. The goal of the proposed scheme is to examine improvements in favor of the user experience and overall satisfaction, while at the same time attempt to propose and implement improvements in favor of the Data Provider (Cloud Provider).

In order to target the maximization of the user's QoE, the proposed scheme uses the Cloud as the base infrastructure to implement and deploy and Context Prediction Algorithm. The Context Prediction Algorithm will predict where the user will be at a given time; it will schedule data delivery to the user and define the triggers which will initiate the delivery of the content. In doing so, this research attempts to maximize parameters that affect QoE by taking into account the user's context. For example and using the scenario above, the proposed research implementation will attempt to have the list of copy centers near the meeting place ready before the user requests it. Operating under the hypothesis that QoE can be maximized by the Cloud's high processing speed, and having data retrieval from the Cloud to the mobile device be performed when cost of the end user is at minimum (WiFi instead of 3G), then the optimization objective can be reached.

III. FURTHER TECHNICAL CONSIDERATIONS

A. Content Prediction

Context Prediction can be useful in order to improve the mobile user's QoE [14]. Algorithms defining methods for context prediction need to be studied and implemented. Such algorithmic approaches that can be utilized include SCP (Structured Context Prediction) [15], Markov Modeling [16] and Spatio-Temporal Data Smoothing Techniques [17].

B. Paraller Data Processing

Following the popularity of the MapReduce programming model and its wide adoption by the industry, we aim in using Hadoop - an open source implementation of MapReduce - as our main data processing structure and model. The goal is to explore the characteristics of a Context Prediction system that can run in parallel.

C. Replication

Since Cloud computations are performed in parallel on a global scale, resources including data and code need to be replicated within the infrastructure. Current research is focusing on replication strategies in Cloud settings that involve optimizing both data processing and network latency [18]. Our twofold approach aims in optimizing processing by accommodating parallel execution via replication, and, moving data to nodes that are closer to the location of the user thus minimizing network latency.

D. Security

Security concerns especially for Public Cloud settings and Context-Aware Applications are amongst the most popular discussion topics in Cloud research. Current Cloud security implementations rely exclusively on the Cloud Provider's capabilities to encrypt data at all processing stages: receiving, storing, transmitting. The cloud user has little to no control of the security mechanisms on a public cloud setting.

Our current research aims in utilizing a user generated security mechanism with user privacy features which will secure user data on the mobile device prior to transmitting it [19]. The research intends to implement security in which the user will be protected even in the case of security breaches of the Cloud Provider.

E. Spatial Filtering

Concepts of Spatial Filtering are intended to be used in our predictive model in an attempt to adapt location to context [20]. Our intent is to deliver data that is filtered by the location of the user, in a context-aware manner, where context considers location as a subset of the contextual information pertaining to a specific mobile user.

IV. DISCUSSION OF KEY ASPECTS

Following the technical considerations for designing and implementing an efficient Mobile-to-Cloud functional architecture, we proceed in stating three important underlying motivations for our work. As our guiding principle, we believe that Cloud Computing can be used efficiently in many technology principles given that its scalability and parallelism capabilities are fully exploited. QoE on the other hand is an important and high-impact subject that directly affects the billions of mobile users around the world. We direct our efforts into improving and optimizing QoE in our venture to support the overall improvement of current technology products.

A. Cloud's Capabilities

This paper adopts the consideration that "computing, storage, and networking [should] focus on the horizontal scalability of virtualized resources rather than on single node performance" [21]. Implementing data processing algorithms which can run in parallel, use cloud-friendly data structures and data storage and take advantage of the Cloud's processing power are essential in realizing the full power of Cloud infrastructure.

B. Mobile User QoE Enhnacements

Research has defined QoE both as a subjective and as an objective measurement. Traditionally, QoE has been measured subjectively via user questionnaires which have been found inefficient and expensive [22]. Objective QoE definitions have been extensively studied using various techniques and attributes in both Application and Network level QoS. These studies have resulted in various correlation models between QoS and QoE [4], [23]. Latest research suggests that QoE can not only be enhanced in its definition using context awareness but also be accurately predicted [5]. For example connectivity – a context aware attribute – can affect QoE. Given that connectivity is directly linked with the user's mobility and mobility has can be accurately predicted, then QoE can be modeled and predicted in a context aware manner. Furthermore, it has been shown that QoE can be used as an optimization metric in order to enhance performance and power consumption of the mobile device [24]. Also, enhancement of QoE can be achieved by benchmarking QoE and dynamically adjusting QoS readings for improving content delivery to the user [25].

Our future QoE centered study will focus in providing a definition for a mobile user's QoE that will include both QoS and context-aware attributes which can potentially be validated via user input. Our aim is to identify and weight all of the attributes can be extracted from the mobile device and the network environment in order to define QoE and use it as a reference for mobile device performance enhancement.

C. Mobile to Cloud Interaction

It has been shown that Mobile to Cloud interactions can be a promising alternative from local data processing and can potentially enhance the QoE of the mobile end user. These interactions have been optimized by adding context information to data relay from the mobile device to the cloud and by scheduling data delivery from the cloud to the mobile device using a context predictive model which keeps communication costs to a minimum. These interaction and cloud resource orchestration needs to be implemented on a public Cloud setting and tested using real world examples.

V. CONCLUSION & FUTURE WORK

This paper proposed, through an illustrative use case scenario, a functional architecture supporting user QoE through context awareness considering the mobile user, as well as the Cloud environment, focusing on the aspect of enabling improved content delivery to a given mobile user employing a computationally expensive application.

The successful content delivery through a functional division of tasks between the mobile device and the Cloud is described by separating the description into the specific functional components comprising the proposed architecture and handling particular aspects of the delivery. Furthermore, additional technological enhancements are considered, i.e. context prediction, parallel data-management, replication, security and spatial filtering.

The innovation of the proposed architecture originates from the fact that it allows optimizations to be enabled at all layers of the content delivery process, i.e. in the receiving layer, the processing layer, the storing layer and the sending layer, making this architecture flexible and efficient. Future work plans to further explore and implement specific enhancements to evaluate their effect on the user QoE on the one hand, through personalized delivery, and the Cloud Provider as a data provider on the other hand, through infrastructure considerations.

REFERENCES

[1] C. Louis, “Gartner Predicts Infrastructure Services Will Accelerate Cloud Computing Growth,” *Forbes*, 2013. [Online]. Available: http://www.forbes.com/sites/louiscolumbus/2013/02/19/gartner-predicts-infrastructure-services-will-accelerate-cloud-computing-growth/.

[2] L. A. Barroso, “Warehouse-scale Computing,” in *Proceedings of the 2010 international conference on Management of data - SIGMOD ’10*, 2010, p. 1000.

[3] F. Agboma and A. Liotta, “Quality of experience management in mobile content delivery systems,” *Telecommun. Syst.*, vol. 49, no. 1, pp. 85–98, Jun. 2010.

[4] K. Rehman Laghari and K. Connelly, “Toward total quality of experience: A QoE model in a communication ecosystem,” *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.

[5] K. Mitra, A. Zaslavsky, C. Ahlund, and C. Åhlund, “Context-Aware QoE Modelling, Measurement and Prediction in Mobile Computing Systems,” *IEEE Trans. Mob. Comput.*, vol. PP, no. 99, pp. 1–1, 2014.

[6] C. Janneteau, J. Simoes, J. Antoniou, M. Kellil, A. Klein, A. Neto, C. Pinto, P. Roux, S. Sargento, and H. D. Schotten, “Context-Aware Multiparty Networking,” in *ICT - MobileSummit Conference*, 2009, pp. 1–11.

[7] J. Antoniou, C. Christophorou, C. Janneteau, M. Kellil, S. Sargento, A. Neto, F. C. Pinto, N. F. Carapeto, and J. Simoes, “Architecture for context-aware multiparty delivery in mobile heterogeneous networks,” *2009 Int. Conf. Ultra Mod. Telecommun. Work.*, pp. 1–6, Oct. 2009.

[8] R. Handorean, R. Sen, G. Hackmann, and G.-C. Roman, “Context aware session management for services in ad hoc networks,” in *2005 IEEE International Conference on Services Computing (SCC’05) Vol-1*, 2005, vol. 1, pp. 113–120 vol.1.

[9] C.-J. Su and C.-Y. Chiang, “Enabling successful Collaboration 2.0: A REST-based Web Service and Web 2.0 technology oriented information platform for collaborative product development,” *Comput. Ind.*, vol. 63, no. 9, pp. 948–959, 2012.

[10] “AWS Elastic Load Balancing – Cloud Network Load Balancing.” [Online]. Available: http://aws.amazon.com/elasticloadbalancing/.

[11] M. Stonebraker, D. Abadi, D. J. DeWitt, S. Madden, E. Paulson, A. Pavlo, and A. Rasin, “MapReduce and parallel DBMSs,” *Commun. ACM*, vol. 53, no. 1, p. 64, Jan. 2010.

[12] C. Douglas, “JSON: The Fat-Free Alternative to XML,” 2006. [Online]. Available: http://www.json.org/fatfree.html.

[13] B. Gil and P. Trezentos, “Impacts of data interchange formats on energy consumption and performance in smartphones,” *Proc. 2011 Work. Open Source Des. Commun. - OSDOC ’11*, p. 1, 2011.

[14] T. Mantoro, A. Olowolayemo, S. O. Olatunji, M. a. Ayu, and A. O. M. Tap, “Extreme learning machine for user location prediction in mobile environment,” *Int. J. Pervasive Comput. Commun.*, vol. 7, no. 2, pp. 162–180, 2011.

[15] S. Zaplata, M. Meiners, and W. Lamersdorf, “Designing future-context-aware dynamic applications with structured context prediction,” no. October 2011, pp. 1185–1204, 2013.

[16] W. Mathew, R. Raposo, and B. Martins, “Predicting future locations with hidden Markov models,” *Proc. 2012 ACM Conf. Ubiquitous Comput. - UbiComp ’12*, p. 911, 2012.

[17] H. Gao, “Mobile Location Prediction in Spatio-Temporal Context,” no. 2, pp. 1–4, 2012.

[18] P. Stuedi and D. Terry, “WhereStore : Location-based Data Storage for Mobile Devices Interacting with the Cloud,” 2010.

[19] M. Mowbray and S. Pearson, “A client-based privacy manager for cloud computing,” *Proc. Fourth Int. ICST Conf. Commun. Syst. Softw. Middlew. - COMSWARE ’09*, p. 1, 2009.

[20] D. Ahlers and S. Boll, “Beyond position-spatial context for mobile information retrieval systems,” in *2009 6th Workshop on Positioning, Navigation and Communication (WPNC’09)*, 2009.

[21] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, “A View of Cloud Computing,” 2010.

[22] H. Gahbiche Msakni and H. Youssef, “Is QoE estimation based on QoS parameters sufficient for video quality assessment?,” *2013 9th Int. Wirel. Commun. Mob. Comput. Conf.*, pp. 538–544, Jul. 2013.

[23] H. J. Kim, D. H. Lee, J. M. Lee, K. H. Lee, W. Lyu, and S. G. Choi, “The QoE Evaluation Method through the QoS-QoE Correlation Model,” *2008 Fourth Int. Conf. Networked Comput. Adv. Inf. Manag.*, vol. 2, pp. 719–725, Sep. 2008.

[24] F. Kaup and D. Hausheer, “Optimizing energy consumption and qoe on mobile devices,” in *2013 21st IEEE International Conference on Network Protocols (ICNP)*, 2013, pp. 1–3.

[25] S. Bischoff, A. Hansson, and B. M. Al-Hashimi, “Applying of Quality of Experience to system optimisation,” in *2013 23rd International Workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS)*, 2013, pp. 91–98.