QoS-Predictions Service: Infrastructural Support for Proactive QoS- and Context-Aware Mobile Services (Position Paper)

Katarzyna Wac^{1,2}, Aart van Halteren², and Dimitri Konstantas^{1,2}

¹ University of Geneva, CUI/ASG, 24 rue General-Dufour, 1211 Geneva, Switzerland

² University of Twente, EWI/ASNA group, P.O.Box 217, 7500 AE Enschede, The Netherlands
{k.e.wac, a.t.vanhalteren, d.konstantas}@utwente.nl

Abstract. Today's mobile data applications aspire to deliver services to a user *anywhere - anytime* while fulfilling his *Quality of Service* (QoS) requirements. However, the success of the service delivery heavily relies on the QoS offered by the underlying networks. As the services operate in a heterogeneous networking environment, we argue that the generic information about the networks' offered-QoS may enable an *anyhow* mobile service delivery based on an intelligent (proactive) selection of 'any' network available in the user's context (location and time).

Towards this direction, we develop a *QoS-predictions service provider*, which includes functionality for the acquisition of generic offered-QoS information and which, via a multidimensional processing and history-based reasoning, will provide predictions of the expected offered-QoS in a reliable and timely manner.

We acquire the generic QoS-information from distributed mobile services' components quantitatively (actively and passively) measuring the application-level QoS, while the reasoning is based on statistical data mining and pattern recognition techniques.

1 Introduction¹

The emergence of new wireless broadband networks and diverse miniaturized and personalized networked devices gives rise to a variety of new mobile services in our daily life. Ultimately, these mobile services are envisaged to be delivered as we move: *anywhere anytime* and under different conditions. However, on the move, these services receive a continuously changing information flow from the environment. A management of this information becomes critical for the delivery of the services. *Context-awareness* seems to be a promising paradigm to manage this flow, where *context* denotes any information that characterizes the user's environment and situation, (e.g. location and time), and any object relevant to the interaction between the user and his service [1].

¹ This is an ongoing research in frame of the Dutch Freeband AWARENESS project (BSIK 03025), partially supported by the INTEROP Network of Excellence (FP6-IST-508011).

A mobile service operates in a heterogeneous networking environment consisting of diverse wireless and wired networks, owned by different providers. As a result, each network is responsible for only a section of the ‘end-to-end’ communication path between a mobile user and an application server, as illustrated in Figure 1. In the case of the *wireless access network*, it can be of diverse technologies (e.g. 2.5G/3G/WLAN), and provided by any provider, including a commercial Mobile Network Operator (MNO).

Fig. 1. An ‘end-to-end’ communication path for a mobile service delivery

Mobile connectivity, that is, persistency of a wireless connection while being mobile, and *Quality of Service (QoS) offered* by this connection are critical factors for the success of the service delivery. This success is measured in terms of fulfilment of the user’s QoS requirements (i.e. *required-QoS*).

However, current rules for the wireless network connectivity are rather simple (and imposed by the MNOs): a default wireless connection, provided by an MNO to which the user is ‘locked-in’, is assumed during the service-design time. As a result, an implicit assumption is made on the network’s offered-QoS. Because a wireless link is usually the bottleneck in an end-to-end communication path, the wireless network’s offered-QoS defines the *end-to-end offered-QoS* and as a result, the fulfilment level of the user’s required-QoS. Furthermore, in today’s mobile services, in spite of the user’s context changes (location, time etc), the required-QoS is assumed to be static during the service delivery.

In our research we take a different direction based on more realistic (in our view) assumptions. The first assumption is that at any location and time various connectivity possibilities (can) coexist and are accessible, offering different QoS to the user. Although this is true today, because of the user ‘lock-in’ restrictions imposed by MNOs, current mobile applications are unaware and unable to make use of these offers. Our second assumption is that the user’s QoS requirements change during the lifetime of the service delivery, depending on the user’s context change. This assumption is valid especially for services that have a long duration, like for example epileptic patients’ continuous mobile health monitoring service ([2], [3]), where the required-QoS will change along the health state of the user. In case of emergency, i.e. epileptic seizure, the required-QoS will be stricter (i.e. shorter delays required), than in case of a non-emergency situation [4]. For other mobile services, the required-QoS can change as, e.g., other mobile services (e.g. mobile TV) overlap with the main service (thus changing its execution environment), or as the user changes roles during the service delivery (as a company’s employee or a father sharing family moments).

For our first assumption - the availability and accessibility of diverse wireless networks at a user’s location and time, we consider not only networks owned by MNOs (2.5/3G) but also other private or public networks (WiFi, WiMax etc). To support the full exploitation of these networks, we make more assumptions. Our third assumption is that a mobile user is empowered to choose amongst the different network providers (and wireless access technologies); he is being able to handover between ‘any’ one of

them, in order to obtain the *best-of the end-to-end offered-QoS* fulfilling his required-QoS². Our fourth assumption is that it is not possible to negotiate end-to-end offered-QoS over the heterogeneous communication path; as a consequence it is possible to obtain only the *best-effort end-to-end offered-QoS*.

With the above four assumptions we overcome the restrictions imposed by the MNOs, and come closer to the reality that the mobile users will face in the future, under the vision of the 4G heterogeneous networking environment [5].

Based on the above assumptions we make two major claims, which define the context of our research. Our first claim is that the mobile service infrastructure must not only be context-aware, but also QoS-aware; aware of the user's required-QoS and the QoS offered by the various networks in user's context (location, time, etc). Our second claim is that the availability of information about these offered-QoS will result in user's intelligent choice of connectivity and an intelligent adaptation of application protocol parameters.

The research concerning our first claim is carried out in the Dutch AWARENESS project [3], and we have already presented it in [4].

In this paper we concentrate on issues related to our second claim. Our starting point is that the availability of run-time generic information regarding the offered QoS by the different available networks may (will, eventually) result in an intelligent (proactive) adaptation of mobile application to the user's environment. From one side, the application will be able to choose 'any' (the best fulfilling required-QoS) suitable network available at his context. From the other side, the application will be able in an intelligent way, to adapt the application protocol according to the network's offered-QoS (for example, by changing the packet size and an inter-packet-delay), in order to obtain best of the best-effort end-to-end offered-QoS. To reach these goals however, we will need *QoS-predictions service provider* as a part of the QoS- and context-aware service infrastructure (or a separate service provider). This service provider will provide generic predictions on the offered-QoS based on historical end-to-end offered-QoS measured and logged by mobile service components (of different users). These predictions will be provided with the use of predictive data mining, pattern recognition techniques, or any other methodology that might be convenient.

A common element for any type of prediction methodology is, nevertheless, the availability of consistent and valid historical data. The focus of this paper is to present our ideas regarding the collection of the end-to-end offered-QoS historical data. Our vision is that the QoS-predictions service provider acquires and aggregates the set of generic end-to-end offered-QoS information that can serve as a basis for predictions for *any* mobile service delivery. The goal is to enable *anywhere anytime anyhow* mobile service delivery obtaining the best of the best-effort end-to-end offered-QoS in an MNO-independent, heterogeneous networking environment, fulfilling the user's required-QoS.

In section 2 we give an overview of the related research, while in section 3 we present the overall system requirements. Section 4 provides our QoS-information model, while section 5 explains information acquisition. The conclusion in section 6 summarizes our research goals and gives an outlook on our research trajectory.

² Further considerations like e.g. the inter-MNO handover speed, or billing, are outside our research scope.

2 State-of-the-Art

This section presents the most important existing research on QoS- and context-awareness as well as on QoS- prediction techniques.

Considering the QoS-awareness aspect, the closest to our goal is the CELLO project, concentrating on a location-based performance evaluation of wireless access networks. Data regarding networks' performance is stored in a GIS system. However, as a network's performance indicator, this project only considers signal strength and not the end-to-end offered-QoS, as we do. Moreover, the overall goal of the project is to enhance the existing mobile operator network [6] and support efficient network extensions; the data is not used for mobile users, as we propose in our research.

Considering the context-awareness, the recent surveys on advances in this research area ([7], [8]) indicate that it focuses mainly on acquiring, processing and using of user context e.g. his location, time, and some physical characteristics of the user's environment and user's personal preferences, based on which the system must adapt. There exists some research in the context-awareness domain on including networks-related context information at the service infrastructure level [9]; however, there is no solution for the reliable network-related context information acquisition. For example, the publication of [10] provides a framework for network-aware applications and indicates that an application-level monitoring is one of the methods for application to be network-aware. However, this publication only considers the end-to-end bandwidth, and no other QoS parameters, as we propose. Moreover, this publication does not provide a solution for information acquisition.

Similarly, [11] and [12] provide an idea on network resource awareness at the application level; both indicate user context as necessary information for an application to adapt. However, both indicate their focus on the wireless access network and not end-to-end resources availability, as we do. Moreover, the mobile connectivity predictions service provider indicated in [12] is based on the single user's history of connectivity, and is used for a user himself to derive further context information. Hence, it will not be used for other users, as we indicate in our research.

In most cases, the research indicates only a signal strength of a wireless access network and this network's theoretical capacity as an indication for offered QoS ([12], [13]). In contrary, we focus on the practical, application-level, end-to-end offered-QoS as the most significant for the mobile service delivery.

There exists some positioning research for context-awareness and prediction techniques using e.g. data mining techniques, however it is very limited - the research presented in [14] focuses on finding an optimal techniques for an application to adapt to the context, rather than to predict the context itself.

Regarding the QoS-awareness of mobile applications, there is a researched solution in which a mobile user of a streaming video application would obtain some information about the offered-QoS directly from the mobile operator network infrastructure [15] e.g. information about his wireless access network performance and location of the video-replica servers. Proposed solution is MNO-specific and does not solve the end-to-end offered-QoS issue, as we do.

Based on this short representation of the ongoing research, we define the innovative contribution of our research to the existing state of the art as the fact that we take the application-level, i.e. end-to-end QoS characteristics as context information.

Moreover, we research the QoS-predictions service provider created based on generic QoS-information obtained by users for users, to support MNO-independent, *anyhow* mobile services delivery obtaining best-of the best-effort end-to-end offered-QoS in a heterogeneous networking environment.

3 System Requirements

3.1 Requirements on a QoS- and Context-Aware Mobile Service Infrastructure

To introduce QoS-awareness into a context-aware service infrastructure, the infrastructure must fulfill a set of requirements (we described them in detail in [4]). Firstly the infrastructure must support user's QoS requirements specification, secondly, the context-aware mapping of user's QoS requirements into the requirements on end-to-end offered-QoS and thirdly, it has to support the service delivery adaptation to the end-to-end offered-QoS, to best fulfill the user's required-QoS. Additionally to these, the context-aware service infrastructure must support *QoS monitoring*; it must support real-time measurements of end-to-end offered-QoS and log those measurements to the QoS-predictions service provider.

3.2 Requirements on the QoS-Predictions Service Provider

As we explained in the introduction, the QoS-predictions service provider is the source of the predictions of end-to-end QoS offered by the underlying heterogeneous networking environment.

The functional requirements on the QoS-predictions service provider are:

- QoS acquisition and storage – the provider must acquire and aggregate results of real-time end-to-end offered-QoS measurements performed by the mobile users, and estimate the quality of the acquired (and aggregated) information
- QoS processing – the provider must transform the acquired data into a form meaningful for inferring further QoS information; inference rules may be dynamically learned along the acquired QoS information
- QoS predictions provisioning - the QoS-predictions provider must provide generic predictions of end-to-end offered-QoS to mobile users, with estimation of the quality of the prediction (i.e. its probability)

The non-functional requirements on QoS-predictions service provider are (after [8], [16]): heterogeneity, interoperability, support for mobility of users, performance, scalability, security and privacy, cost, traceability and control and fault tolerance.

4 QoS-Information Model

In order to support the QoS-predictions, measured QoS information as well as other context information must be logged. It is thus of major importance to define in details what type of data are needed to develop a generic QoS-predictions service provider. In the following sections we describe in detail structure of the logged generic end-to-end

offered-QoS information, and namely a) end-to-end QoS information (section 4.1) and associated context information (section 4.2). All the generic QoS-information needs to be acquired, processed and provided along with its quality measures like e.g. information freshness and accuracy. These quality measures are denoted as a *Quality of Context (QoC)* and they play a significant role in QoS-information processing phase [17].

4.1 QoS-Information Model

Following the classification given in [18] we focus on a generic *network-oriented* and selected *user-oriented* end-to-end offered-QoS information that our QoS-predictions service provider acquires, aggregates, processes and provides:

- network-oriented end-to-end offered-QoS information is the end-to-end communication path's performance [16] expressed in terms of:
 - speed - end-to-end application-level data delay and goodput³
 - accuracy - data corruption probability
 - dependability - service availability, service reliability in terms of application-level data loss probability
- user-oriented end-to-end offered-QoS information is:
 - service cost, e.g., per use or per unit cost
 - security level, e.g., authentication, confidentiality, integrity, non-repudiation

The user-oriented perceived QoS, e.g., picture resolution, video rate/smoothness, audio quality, AV synchronization, are outside of the scope of our QoS-predictions provider due to its application-domain dependency. Our vision is that the QoS-predictions service provider deals with the set of *generic* end-to-end offered-QoS information that can serve as a basis for *any* mobile service delivery adaptation.

4.2 Associated Context-Information Model

The network- and selected user-oriented end-to-end offered-QoS information described in the previous section needs to be acquired from, and disseminated to mobile users along with the following service usage's generic context information (non exhaustive list):

- Mobile user's geographical location and time – where (and when) the end-to-end offered QoS was measured or for where (and when) the QoS-predictions are given
- Mobile user's mobility level – velocity of changes of user location in time
- Wireless access network provider used e.g. Vodafone, and wireless access network technology e.g. 3G; associated network's usage cost and security level
- Mobile service used, application-protocol and its traffic model type – we define 1) an interactive, 2) messaging and 3) streaming traffic models described in terms of a) packet size distribution and b) inter-packet-delay distribution

³ Goodput is an application-level throughput.

- Transport-level protocol, e.g. TCP or UDP and its parameters
- Mobile device used and its technical specifications like, e.g. OS and its version, CPU and its load level, memory type/size and its usage level, battery type/size and its level, wireless access networks interfaces supported
- Application server used and its technical specifications like, e.g. OS and its version, CPU and its load level, memory type/size and its usage level

Other type of context information can be included, depending on the specific needs of the mobile application and the prediction service provider.

5 End-to-End Offered-QoS Information Acquisition

The QoS-predictions' service provider acquires and accumulates logs on end-to-end offered-QoS measured by mobile service components of different users. Particularly, the results of run-time quantitative measurements of end-to-end application-level offered-QoS are logged. This can be done in two ways: first, actively by a standalone *evaluation system* – distributed mobile application dedicated in performing the measurements (section 5.1) and second, passively during the execution of a mobile application (section 5.2)

5.1 Active End-to-End Offered-QoS Measurements – The *Evaluation System*

To actively measure the end-to-end QoS offered by the underlying heterogeneous networking environment, we first considered the existing performance evaluation tools as presented in [18]. However, as these tools are mainly fixed-networks oriented they do not fulfill the basic mobility, heterogeneity and distribution requirements. For this reason we have developed our own a scalable distributed non-invasive measurements-based performance evaluation system (described in [19]) consisting of:

- *workload generator* – for generating a controlled application-level workload (i.e. along the application-protocol traffic model type) at ingress points of the heterogeneous networking environment;
- *measurement function* – for acquiring and correlating measurements (i.e. application-level's events-based timestamps), at the ingress and egress points of the heterogeneous networking environment and for generating statistically analyzed end-to-end measurements results in a textual and graphical form.

Fig. 2. The evaluation system in the 'end-to-end' communication path

Figure 2 presents the distribution of evaluation system in the end-to-end communication path. The critical requirement for this setup is that the distributed nodes have their time clocks precisely synchronized (by e.g. GPS) in order to get reliable timestamps.

Our evaluation system measures speed-related performance charac-

teristics - one-way end-to-end delays and goodput along with the user context information (section 4.2) e.g. user location, time, wireless access network provider/technology and technicalities of mobile device/application server used. The current implementation of our system used only TCP as a transport protocol and we are expanding it towards use of the UDP protocol.

The system has been used in the measurement of end-to-end offered-QoS for a continuous mobile health monitoring application - the *MobiHealth* application [2] operational in a heterogeneous networking environment with a 3G-UMTS wireless network as a wireless access network, as presented in [20] and [21].

5.2 Application-Level Passive End-to-End Offered-QoS Measurements

To passively measure generic end-to-end QoS offered by the underlying heterogeneous networking environment, we are now working on an instrumentation of our *MobiHealth* application, developed and operational at the University of Twente, the Netherlands. The application itself acts as a workload generator, and we are currently working on the implementation of a generic measurement function within it.

6 Conclusions and Outlook

In this paper we discussed our ongoing research on QoS- and context-aware service infrastructure that supports the development of mobile applications in a heterogeneous networking environment. Particularly we presented our vision of *anywhere anytime anyhow* mobile service delivery fulfilling the user's required-QoS by using the generic predictions of end-to-end offered-QoS in user's actual context (e.g. his location and time) to obtain the best-of the best-effort end-to-end offered-QoS. We presented issues related to the collection of generic QoS measurements.

The main functionality of the QoS-predictions service provider is the processing of the logged generic QoS-information (i.e. end-to-end offered-QoS) and the provision of generic predictions of the end-to-end offered-QoS to mobile applications. We envision that the QoS-information processing is based on the data mining and patterns recognition techniques. We anticipate that the Quality of Context information associated with the QoS-information will play an important role in the QoS-information processing. However, we have not acquired any concrete results yet as the research is still ongoing.

In our research, the mobile entities are generic wireless sensor/actuator network nodes - logging the generic QoS-information to the QoS-predictions provider and acting upon the generic QoS-predictions, hence changing the heterogeneous networking environment's offered-QoS.

We identify the novelty of our approach as being the expansion beyond the standard QoS management framework, comprising of QoS contracts and QoS negotiation components. We propose a complete user-driven approach, where a mobile user (and particularly an application on the user's behalf) can make an intelligent choice amongst the underlying network providers and technologies to best fulfill the mobile user's required-QoS. We indicate the importance of our eventual findings and its applicability to fulfill required-QoS for any mobile service user, in an *anyhow* manner.

The target audience of our research is any mobile service provider. As more and more applications go mobile, the demand for a generic QoS-information provided in a reliable and timely manner, will be growing. The beneficiaries of our research will be directly mobile users, experiencing their mobile services at the required quality level.

Our research expands beyond the current telecom business model, where the user is locked to one MNO and technologies offered by this operator. In our view, the user should be empowered to make a decision on which network technology offered by which provider is the most convenient for him to use. Moreover, by introducing the generic QoS-predictions service provider we further indicate the user-empowerment; generic QoS-information is acquired by users and it is used exclusively for the users.

References

1. Dey, A. The Context Toolkit: Aiding the Development of Context-Aware Applications. In Workshop on Software Engineering for Wearable and Pervasive Computing. 2000. Limerick, Ireland.
2. van Halteren, A., et al., Mobile Patient Monitoring: The MobiHealth System. The Journal on Information Technology in Healthcare, 2004, 2(5): p. 365-373.
3. Wegdam, M. AWARENESS: A project on Context AWARE mobile Networks and Services. In 14th Mobile & Wireless Communication Summit. 2005. Dresden, Germany.
4. Wac, K. Towards QoS-Awareness of Context-Aware Mobile Applications and Services. In On the Move to Meaningful Internet Systems 2005: OTM Workshops, Ph.D. Student Symposium. 2005. Agia Napa, Cyprus: Springer Berlin / Heidelberg.
5. Frattasi, S., Fathi, and F. Fitzek, 4G: A User-Centric System. Wireless Personal Communications Journal (WPC) - Special Issue on Advances in Wireless Communications: Enabling Technologies for 4G, 2006.
6. Horsmanheimo, S., H. Jormakka, and J. Lähteenmäki, Location-Aided Planning in Mobile Networks - Trial Results. Wireless Personal Communications: Special Issue on Cellular and Wireless Location Based Technologies and Services, 2004, 30(2-4): p. 207-216.
7. Chen, G. and D. Kotz, A survey of context-aware mobile computing research, in Technical Report TR2000-381. 2000, Dept. of Computer Science, Dartmouth College.
8. Henriksen, K., et al. Middleware for Distributed Context-Aware Systems. In On the Move to Meaningful Internet Systems 2005. 2005. Agia Napa, Cyprus: Springer Berlin / Heidelberg.
9. Gloss, B. System Architecture of a Mobile Message Transport System. In 11th Open European Summer School - Networked Applications (EUNICE05). 2005. Madrid, Spain.
10. Bolliger, J. and T. Gross, A framework based approach to the development of network aware applications. IEEE Transactions on Software Engineering, 1998, 24(5): p. 376-390.
11. Chalmers, D. and M. Sloman. QoS and Context Awareness for Mobile Computing. In 1st Intl. Symposium on Handheld and Ubiquitous Computing (HUC99). 1999. Karlsruhe, Germany: Springer.
12. Sun, J., J. Sauvola, and J. Riekk. Application of connectivity information for context interpretation and derivation. In 8th International Conference on Telecommunications (ConTEL05). 2005. Zagreb, Croatia.
13. Peddemors, A., H. Eertink, and I. Niemegeers. Communication Context for Adaptive Mobile Applications. In 3rd IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW05). 2005.

14. Mähönen, P., et al. Cognitive Wireless Networks: Your Network Just Became a Teenager. In INFOCOM06. 2006. Barcelona, Spain: IEEE.
15. Han, Q. and N. Venkatasubramanian, Information Collection Services for QoS-aware Mobile Applications. *IEEE Transactions on Mobile Computing*, 2006. **5**(5): p. 518-535.
16. ITU-T, General aspects of Quality of Service and Network Performance in Digital Networks, including ISDNs. 1993, ITU.
17. Widya, I., B.J. van Beijnum, and A. Salden. QoC-based optimization of end-to end m-health data delivery services. In 14th IEEE International Workshop on Quality of Service (IWQoS06). 2006. New Haven CT, USA: IEEE Press.
18. Michaut, F. and F. Lepage, Application-oriented network metrology: Metrics and active measurement tools. *IEEE Communications Surveys & Tutorials*, 2005. **7**(2): p. 2-24.
19. Bults, R., A. van Halteren, and K. Wac, Design of a Scalable Distributed End-to-End Performance Evaluation System. 2005, University of Twente, the Netherlands.
20. Wac, K., et al. Measurements-based performance evaluation of 3G wireless networks supporting m-health services. In 12th Multimedia Computing and Networking (MMCN05). 2005. San Jose, CA, USA.
21. Bults, R., et al. Goodput Analysis of 3G wireless networks supporting m-health services. In 8th International Conference on Telecommunications (ConTEL05). 2005. Zagreb, Croatia.