# Quality Driven Web Services in Mobile Computing

M. Younas[*], K-M Chao[*], N. Griffiths[♠], R. Anane[*], I. Awan[♣]
[*]*School of MIS, Coventry University, Coventry, UK*
*{m.younas, k.chao, r.anane}@coventry.ac.uk*
[♠]*Department of Computer Science, University of Warwick, UK*
*nathan@dcs.warwick.ac.uk*
[♣]*Department of Computing, University of Bradford, Bradford, UK*
*i.awan@scm.brad.ac.uk*

## Abstract

*Recent advances in Web and mobile computing technologies provide users with a vast amount of information and services such as online news, shopping, music, movies, and video games. These services can be provided with different qualities in terms of response time, price to pay, resource type (e.g. PDA or mobile phone), and audio and video reproduction quality. These factors must be considered in order to achieve an acceptable level of quality of the required services. However, currently there are no standard criteria to assess the quality of such services. In this paper, we propose novel agent-based criteria that includes methods for assessing the quality of services and resources. These criteria, based on supply chain methods, enable agents to select and execute requests appropriately, depending on the quality of service and the underlying system's capability. The proposed approach is generic and can be applied to different domains such as e-commerce, m-commerce, and multimedia applications.*

## 1. Introduction

With recent technological developments of wireless communication networks, handheld devices (e.g. PDAs), and software standards (e.g. J2ME), the area of mobile computing is becoming increasingly important. Mobile computing enables people to access information and services in an anywhere and anytime fashion [1]. For example, using a mobile phone and handheld PC, a user can access a variety of information and services ranging from text messages to Web content and mobile commerce services [2, 3].

Currently there is growing trend to access and manipulate Web services using mobile computing technology [4, 5, 6, 7]. Web services have recently gained enormous popularity both in industry and the research community. Web services enable the development of various applications using standard technologies and protocols such as XML, SOAP, WSDL, and UDDI. Web services provide a standard means of communication and integration among different distributed systems and applications running on a variety of platforms. For example, different specialist music stores can be integrated through Web services in order to provide customers with a set of integrated facilities.

Within mobile computing environments Web services are called mobile or M-services [6, 7]. In the following discussion we use the term 'M-services' or simply 'services' to refer to the use of Web services in mobile computing.

The work carried out in this paper is motivated by issues surrounding the quality of M-services. Many factors contribute to the complexity of assessing the quality of such services. For instance, M-services can be provided by different independent service providers with different qualities according to such factors as response time, price to pay, complexity, availability, and device compatibility. Mobile devices are generally resource-scarce as they have limited memory, computational resources, and battery power. In addition, they have limited bandwidth and less reliable wireless connectivity when compared to traditional fixed computing devices. These characteristics complicate the process of assessing and maintaining the quality of M-services.

Currently, there are no standard criteria to assess the quality of M-services. However, some interesting approaches are emerging in this direction. Maamar et. al. present a framework for the provision, composition, and selection of Web services in mobile as well as fixed computing environments [6, 7]. In this framework software agents are used in order to represent and work on behalf of users, service and resource providers. The framework aims to optimally select fixed and mobile computing resources that adapt to the requirements of Web services (which will use those resources). This framework is valuable, however, it is limited to considering cost and reliability factors in the selection of required services and resources.

In this paper, we propose agent-based criteria that include methods for dynamically assessing the quality of services. These criteria, based on supply chain methods, enable agents to select and execute the most appropriate requests, depending on the quality of service and the underlying system's capability, such as the nature of the mobile devices and of the network itself. The proposed criteria take into account multiple factors such as tolerance, substitution, generality, complexity, and processing time of the services. We see that the flexibility, adaptability, and high-level semantics afforded by multi-agents provide a useful framework for assessing the quality of services. Agents can function correctly and make decisions even the processing unit (such as a mobile device) is disconnected from the network [8]. Moreover, agents can perform complex tasks and autonomous interaction when connectivity is present.

The remainder of this paper is structured as follows. Section 2 provides background information in order to establish basic concepts, definitions, and related issues. In Section 3 we present the proposed criteria for assessing the quality of M-services and describe how agents can utilize the proposed criteria in quality assessment. Finally, Section 4 concludes the paper and identifies future research work.

## 2. Background

This section provides background information including an overview of mobile computing architectures, Web services, agent technology, and the characteristics of Web services within mobile computing environment.

### 2.1 Mobile computing

A generalized architecture of mobile computing is shown in Figure 1 [1, 2]. This architecture contains different components: *fixed-hosts* (FH), *base stations* (BS) or *mobile support stations* (MSS) — these are computer systems, which are inter-connected via a wired network. BSs are capable of communicating with mobile units through a wireless interface.

Mobile units are portable computers or WAP-enabled mobile phones. BSs also act as an interface between mobile units and systems of the wired-network. Operations of multiple BSs are coordinated by a *Base Station Controller* (BSC). BSCs are in turn controlled by the *Mobile Switching Center* (MSC), which is connected to the *Public Switching Telephone Network* (PTSN). Each BS covers a particular area, called a *cell*. Mobile units can move from one cell to another. This process of moving is referred to as *handoff*.

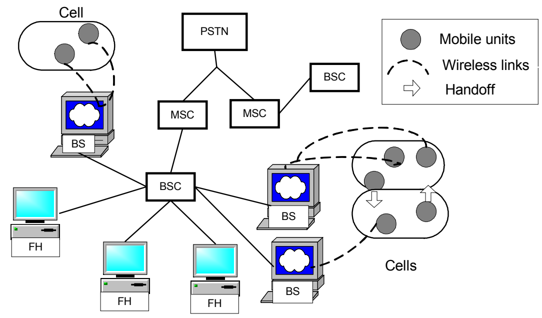


Figure 1: An Architecture of Mobile Computing

### 2.2 Web services

Web services enable the development of diverse applications using standard technologies and protocols such as Extensible Markup Language (XML), Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI) [4]. These technologies and protocols are organized into different layers comprising: network, messaging, service description, service publication and service discovery layer. Web services use commonly deployed network protocols such as TCP/IP, HTTP, and IIOP. Web Services provide a standard means of integration and communication among different distributed systems and applications running on a variety of platforms. For example, different music stores can be integrated through Web services in order to provide customers with particular services.

In mobile computing environments, Maamar *et. al.* [6, 7] define Web services in two different ways: (i) Web services are called mobile or M-services if they are remotely activated using a mobile device, or (ii) Web services are M-services if a Web service can be transferred from host system to a mobile device where its execution takes place. In [6, 7] the above definitions (i) and (ii) are respectively termed as weak and strong definitions of M-services.

### 2.3 Agents

In the proposed approach we represent different services, resources, and users as agents. However, in order for such representations to be effective the agents must exhibit certain characteristics, i.e. they must be autonomous, pro-active, responsive to their environment, and be able to socially interact with others. We adopt a BDI-based approach of taking an agent to comprise: beliefs about itself, others and the environment; a set of desires (or goals) representing the states it wants to achieve; and intentions corresponding to the plans adopted in pursuit of these desires [9, 10, 12]. Agents have a library of partial plans from which to select. These plans represent how various services and resources are combined to provide useful facilities to users. An agent responds to changes in its beliefs, resulting from perception, by generating goals. These goals are then evaluated, and the most important are adopted as intentions by selecting an appropriate plan from a set of applicable plans, and committing to its execution. Finally, an agent selects a particular intention to pursue and acts toward its achievement. Several alternative agent architectures are possible, and the criteria that we describe in the following sections can be applied equally to these alternatives.

### 2.4 Characteristics of M-services

In this section we illustrate the characteristics of M-services in the architecture shown earlier in Figure 1.

**Low Reliability and Availability:** Mobile units are frequently unavailable due to many reasons. Firstly, they have limited power supplies. This creates serious issues for M-services. For example, a mobile unit having a low battery may not communicate with other units involved in the processing of a service. Furthermore, the mobile unit may be stolen, or it may be affected by break down, and so forth. All such events affect the functionality of mobile units, and consequently the respective services. Thus, it is unrealistic to assume that a mobile unit is always available to participate in processing an M-service.

**Limited Bandwidth:** Current mobile technology is associated with a limited bandwidth. Although the bandwidth is increasing, it is assumed that it will remain limited compared to the wired network. However, M-services may need the transfer of large amounts of data and involve complex processing [11]. Furthermore, service discovery (as through UDDI) requires multiple messages to be communicated between the mobile device and the service providers. The consequence is that communication links may be overloaded with a high volume of information exchange and they may not be able to transfer the required information to the mobile users.

**Limited Storage:** In addition to bandwidth issues, the lower disk and memory resources of mobile devices are not capable of handling large volumes of data or processing complex applications as in M-services. For example, a user with a handheld device might use a service that produces hundreds of inventory records. Such data and applications cannot be managed and stored directly using a handheld PC.

**Complexity of Mobility:** In mobile computing, data, applications, and also the processing units (such as PDAs or mobile phones) move. This mobility greatly complicates the processing of a service. For example, an application that sends a request for a service may move from one location to another as their originating devices move. Thus, they may originate at one site and terminate at other site. Furthermore, the handoff process also creates complexity in M-services, as handoff is generally unpredictable. Thus, it may affect the termination process of an application. This is because the location of the desired mobile device associated with a service may not be immediately available for communication.

## 3. Quality assessment criteria

We propose a set of criteria for agents to assess the quality of services in mobile computing environments. Users are represented by agents who assess the quality of potential service providers in fulfilling user requests. Furthermore, service providers are represented by agents who make similar assessments when performing service composition to provide complex composite services. These criteria are devised using supply chain techniques [2]. The basic element in these criteria is to determine the degree of *availability* of services. The degree of availability is an objective measure and is determined by inspecting knowledge of plans and the current situation. This approach is a generic and novel decision making technique. It includes multiple attributes to enable the agent to assess the overall availability of services over multiple factors comprising: tolerance, substitution, generality, processing time, and complexity. This technique determines the individual degree of availability for each factor, and sums weighted degrees of availability in order to produce an overall measure. This is used by an agent in order to evaluate each possible plan for completing a request or providing a service in order to select the best option. We use the example application of an online music store in order to illustrate the degree of availability. We compute the individual degree of availability of each factor as probabilities that range over the interval [0,1]. This is because exact values of these factors vary depending on the nature of the mobile devices, applications and the operational environment.

**Tolerance**: Tolerance is defined as the degree of precision that a particular service contained in a plan successfully completes. Tolerance is computed below using statistical measurements.

Suppose a mobile application *A* depends on services $S_1$, $S_2$, and $S_3$ with increasing precision requirements $1\sigma$, $2\sigma$, and $3\sigma$, respectively. The degree of tolerance for application *A* is calculated as follows.

$$\beta_l = 1 - \frac{(S_1 + S_2 + S_3)}{3} \qquad (1)$$

In the above equation, $S_1$, $S_2$, and $S_3$ are computed as follows.

$$S_1: \overline{X}_{S1} \pm 1\sigma = 73.27\%$$
$$S_2: \overline{X}_{S2} \pm 2\sigma = 95.45\%$$
$$S_3: \overline{X}_{S3} \pm 3\sigma = 99.73\%$$

where $\sigma = \frac{\sqrt{(Xi-\overline{X})^2}}{i}$; a standard deviation.

The tolerance level of $A$ is

$$\beta_1 = 1- (73.27\% + 95.45\% + 99.73\%)/3$$

Suppose, for example, that a user downloads an audio file onto his/her mobile device using the Web service of the music store. Let us assume that the total download time is 60 seconds. If the network connection is reasonably free then it will take 45 seconds in the download process. Similarly, an estimated time for the mobile device to process the download request (for the audio file) is 50 seconds. If the network connection is available for the overall duration, then the song can be downloaded within the required time. However, if the network is busy and it takes 15 extra seconds, then the required song cannot be downloaded within the specified time. In this case, the precision of the service time is 73.27% for network and 95.45% for the mobile device. Thus the required tolerance level for the application $A$ to be processed in time is $\beta_1 = 1 - 84.36\%$.

**Substitution**: If a particular service $S$ (e.g. a Britney Spears song) can be replaced with alternative services $S_1$, $S_2$, and $S_3$ (e.g. Back Street Boys songs), then degree of availability of service $S$ is high. A higher number of substitutes implies a higher the degree of availability. This is computed as follows

$$\beta_2=1-(1-X_{S1}) \times (1-X_{S2}) \times (1-X_{S3}) \qquad (2)$$

where $X_{Si}$ represents the compatibility of substitutes with service S.

Suppose, in our example, that the music store has contracts with a number of other stores so as to arrange for alternative songs for a customer if their original choice is unavailable for some reason. Arranging for alternative songs may result from excessive user demand for a particular song. Suppose that the user can be provided with three alternative songs, and let the possibly of getting the song from service $S_1$ be $X_{S1} = 0.9$, and not getting the song be $1- X_{S1}$. The same principle can be applied to services $S_2$ and $S_3$ such that $X_{S2} = 0.8$ and $X_{S3} = 0.85$. Thus the existence of substitutes is $\beta_2 = 97\%$.

**Generality**: Service $S$ can be considered generic if it can be provided by a number of service providers. The degree of availability of $S$ can be calculated as

$$\beta_3=1-(1-X_{S1}) \times (1-X_{S2}) \times (1-X_{S3}) \times (1-X_{Sk})$$

where $(1-Xi)$ represents the possible shortage of service $S$ by a service provider $i$.

For example, different movies (or video files) can be provided by different music stores (i.e. service providers). If a movie $m_1$ is not available at one store, then the degree of availability depends on how many other stores can provide the same, or an equivalent, movie. Thus $\beta_3$ is calculated as

$$\beta_3 = 1-(\beta_2 \times (1-X_{S1}) \times (1-X_{S2}) \times (1-X_{S3})) \qquad (3)$$

If the estimated values of $X_1 = 0.8$, $X_2 = 0.9$ and $X_3 = 0.95$, then $\beta_3 = 99.997\%$. Thus, the more generality the higher the degree of availability. Note that *generality* is related to the *substitution* as shown in equation (3).

**Processing Time**: The importance of a service $S$, with respect to a plan $P$, in terms of processing time can be evaluated as

$$\beta_4 = 1- \frac{t}{t_{total}} \qquad (4)$$

where $t$ and $t_{total}$ denote the required slack processing time for service $S$, and the total slack processing time for plan $P$, respectively.

For example, suppose that the overall download slack time, $t_{total}$, (for a song) is 60 seconds and the network slack time, $t$, is 20 seconds. The importance of network speed to the song download in terms of processing time is 33.33 % and so $\beta_4$ is 67.77%. Thus, the higher the processing time percentage the lower the degree of availability.

**Level of Complexity**: The level of complexity is computed by considering the number of processing steps $v$ and the number of related components $h$ in the plan representing a particular service $S$. This is calculated as follows.

$$C_S = 1-(v \times h)$$

So, the significance of service $S$ to the plan $P$ can be formulated as

$$\beta_5 = -1 \times \frac{C_S}{C} \qquad (5)$$

where $C_S$ is the overall complexity of plan $P$.

The above equation (5) states that the higher complexity the lower the degree of availability.

For example, to successfully download a movie clip a user may need specialised software. It is not assumed that such software is pre-installed on the user's machine. For example, suppose that the probability of successfully downloading the required movie is 90% and the probability of successfully downloading the specialised software is 85%. With the requirement of specialised software, the probability of downloading the movie is 1-(90% × 90% × 85%) = 31.15%. If the specialised software is not required then the probability of downloading the movie is reduced to 19%. If the overall complexity is 200%, then $\beta_5$ = -15.50% (with specialised software) and $\beta_5$ = -9.5% (without specialised software).

**Reliability**: The degree of availability of a service depends on its reliability (denoted as $\beta_6$). The evaluation of service *A* can be carried out by inspecting its previous performance. If the success rate of service *S* falls within the range of $\overline{X}_A \pm 1\sigma$, its reliability is 73.27%. If there is no previous data, the default value, $\beta_6$, will be derived from $\overline{\beta}_6$. The value of its reliability is revised after each request.

**Overall Availability:** In the above we have calculated different factors for assessing the quality of services. Now we calculate the overall degree of availability which is based on the above factors of tolerance ($\beta_1$), substitution ($\beta_2$), generality ($\beta_3$), processing time ($\beta_4$), the level of complexity ($\beta_5$) and the reliability ($\beta_6$). The overall degree of availability $D_{availability}$ is determined as follows:

$$D_{availability} = (w_1 \times \beta_1) + (w_2 \times \beta_2) + \ldots + (w_k \times \beta_k)$$

$$D_{availability} = \sum_{i=1}^{k} W_i \times \beta_i \qquad (6)$$

where $w_k$ denotes the weights for factor *i* and $\sum_{i=1}^{k} w_i = 1$.

The resulting value, $D_{availability}$, represents the availability of the service *S*. The higher the value of $D_{availability}$, the better the availability of the service is. Agents can utilise this to evaluate each possible intention or plan in order to select the best option according to the nature of services and resources given the characteristics of the mobile devices. This assessment is made both by a user's agent in fulfilling a request, and a service provider in performing service composition.

If a plan is made up of a number of sub-plans *n* and requires service composition, then for a service $S_1$, the degree of availability is determined as follows:

$$D_{availability\text{-}n} = \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \beta_{ij} * W_{ij}\right) \qquad (7)$$

Agents use the above criteria to plan and execute user requests according to the nature of the requested service and the characteristics of the mobile devices. In particular, these criteria can be seen as complimenting traditional planning heuristics, such as cost and length, to allow an agent to select an appropriate plan to achieve its goals. We represent both users and service providers by BDI-based agents, and this mechanism allows these availability criteria to be utilized both in terms of a user's agent selecting a service to fulfill a request and a service provider instantiating the service. Thus, the applicable plans are not only a set of unordered pattern or function matched plans, but also an ordered set of plans based on the quality ratings to enable the agent to select. In addition, the agent can introduce these criteria to filter the plans in order to eliminate low quality plans from being included in the applicable plans. So, the plans with or above a certain degree of availability will form the set of applicable plans from which the agent can select. After a plan has been carried out, feedback can be incorporated into the system in order to reflect its up-to-date availability.

## 6. Conclusion and future work

In this paper we have presented an agent-based approach for assessing the quality of Web services in mobile computing environment. We have defined a set of criteria that take into account multiple factors in order to assess the quality of services and the respective resources. These factors, based on supply chain techniques, provide an objective mechanism to assess the availability of a service in terms of tolerance, substitution, generality, processing time, complexity and reliability. Assessments of service availability can be used by agents, representing users, services and resources, to ensure that an appropriate service quality is achieved. This allows agents to factor characteristics of the application environment, such as limited bandwidth and storage, into their reasoning. The proposed criteria are generic and can be applied to various domains depending the nature of applications. For example, these criteria are applicable to mobile or M-commerce applications in order to assess the quality of required services as indicated above.

The main area of future work is to incorporate the notion of trust into an agent's assessment of service. Trust provides a means of assessing the perceived risk of interacting with a particular agent [13]. Availability can be seen as an objective measure, while trust is a subjective measure based on an agent's individual experience.

The integration of an emerging Web Service ontology standard, DAML-S [14], and intelligent agent technologies [15] attempts to provide a way to facilitate the automation of Web services composition [15, 16, 17]. The issue of rating the quality of Web services for the selection of a service is recognized by the DAML services coalition as an important factor by introducing QualityRating attribute in the service profile. However, no clear guideline is currently given to derive the rates of quality. The work we are currently undertaking is to formulate the proposed criteria along with aforementioned notions of risk and trust with DAML+OIL [18]. This will provide a subset of semantic markup within DAML-S in order to provide a means for evaluating the quality of web services.

## References


[1] A. Helal, B. Haskell, J.L. Carter, R. Brice, D. Woelk, M. Rusinkiewicz *"Any Time, Anywhere Computing: Mobile Computing Concepts and Technology"* Kluwer Publisher, 1999

[2] M. Younas, K-M Chao, R. Anane "M-Commerce Transaction Management with Multi-Agent Support" *17th International Conference on Advanced Information Networking and Applications (AINA)*, IEEE CS Press, March 27-29, 2003, Xi'an, China

[3] K. Kuramitsu, K. Sakamura " Towards Ubiquitous Database in Mobile Commerce" 2nd ACM *International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE01),* May 20, 2001 in Santa Barbara, California, USA

[4] "Web Services Activity" W3C http://www.w3.org/2002/ws/

[5] "Java 2 Platform, Micro Edition (J2ME)" Sun Microsystems, Inc. http://java.sun.com/j2me/

[6] Z. Maamar, Q.Z. Sheng, B. Benatallah "Selection of Web Services for Composition Using Location of Provider Hosts Criterion" *Ubiqitous Mobile Information and Collabration Systems Workshop* (UMICS 2003), held with CAiSE'03, June 2003, Austria.

[7] Z. Maamar, Q.Z. Sheng, B. Benatallah "On Composite Web Services Provisioning in an Environment of Fixed and Mobile Computing Resources" *Information Technology and Management,* Kulwar Academic Publishers, Vol 5, No 3, 2004 (forthcoming)

[8] S.W. Loke "Supporting Intelligent BDI Agents on Resource-Limited Mobile Devices – Issues and Challenges from a Mobile Database Perspective" *Autonomous Agents & Multiagent Systems (AAMAS) Workshop*, July 2002, Bologna, Italy.

[9] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice," Knowledge Engineering Review, vol. 10, no. 2, pp. 115–152, 1995.

[10] M. E. Bratman, D. Israel, and M. Pollack, "Plans and resource-bounded practical reasoning," Computational Intelligence, vol. 4, pp. 349–355, 1988.

[11] Adrian Mello Can Web services drive mobile apps? May 06, 2002, ZDNet Australia http://www.zdnet.com.au/newstech/ebusiness/story/0,2000048590,20265026,00.htm

[12] M. Luck and M. d'Inverno, "A formal framework for agency and autonomy," in Proceedings of the First International Conference on Multi-Agent Systems. AAAI Press/The MIT Press, 1995, pp. 254–260.

[13] C. Castelfranci and R. Falcone, "Principles of trust for MAS: Cognitive anatomy, social importance, and quantification", in Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98), pp. 72-79, Paris, France, 1998.

[14] The DAML Services Coalition, http://www.daml.org/services/daml-s/0.9/

[15] Paul A. Buhler, Jose M. Vidal, Semantic Web Services as Agent Behaviors, *Agentcities: Challenges in Open Agent Environments*. Springer-Verlag, 2003, pp 25-31

[16] Debbie Richards, Sander van Splunter, Frances M.T. Brazier, Marta Sabou, "Composing Web Services using an Agent Factory", *AAMAS Workshop on Web Services and Agent-Based Engineering,* 2003.

[17] McIlraith, S., and Son, T. C. "Adapting golog for composition of semantic web services", *Conference Proceedings of.* KRR, pp 482-493, 2002.

[18] The DAML joint committee, http://www.daml.org