Paper presented at the IST Mobile Summit

Galway, October 2000

Performance Evaluation of the WAP Protocol over SMS in a GSM Network

Alessandro Andreadis[1], Giuliano Benelli[1], Giovanni Giambene[1], Bernardo Marzucchi[1], David Sennati[1]

[1] Dipartimento di Ingegneria dell'Informazione - Università degli Studi di Siena, Via Roma, 56 - 53100 Siena, Italy, email: benelli@unisi.it

Abstract: *The widespread diffusion of mobile communication systems pushes research and development efforts towards the implementation of new services accessible through mobile terminals. This work has been carried out in the framework of the PALIO Project in the fifth Research Framework of the European Commission. It has been considered a tourist service where each user may access information directly through its mobile GSM phone by means of the Wireless Application Protocol (WAP). This paper investigates the performance of WAP over the short message service. A suitable traffic model has been identified and a GSM signalling configuration has been selected to provide the information service at the city level. Accordingly it has been possible to evaluate the delay required to transmit WAP decks to users. This study is particularly interesting because it provides a tool for sizing the envisaged tourist service.*

1. Introduction

At present mobile communication systems are faced with the need to provide a wireless access to the Internet. A recently arisen solution is given by the *Wireless Application Protocol* (WAP) [1],[2] that allows a mobile terminal with a little display to browse Web pages or suitably designed pages in the *Wireless Markup Language* (WML). The WAP protocol is air interface independent and allows the easy integration of multimedia services and customised applications. WAP contains the definition of a microbrowser according to which WML and WMLScript are interpreted in the handset and presented to the user. Since a mobile user can not use a QWERTY keyboard or a mouse, WML documents are structured into a set of well-defined units of user interactions called *cards*. A single collection of cards is called *deck*, which is the unit of content transmission and is identified by a *Uniform Resource Locator* (URL). After loading a deck, a user agent displays the first card; then, the user decides whether to proceed or not to the next card of the same deck. Some critical aspects that need to be investigated for the adoption of WAP are as follows:

- The robustness of the WAP protocol in the presence of frequent packet errors.
- Integration of multimedia services over WAP.
- The achievable *Quality of Service* (QoS) of applications over WAP (e.g., the prompt delivery of the requested WML decks) depending on the bearer services used in the mobile network.

This paper focuses on the last aspect and assumes the use of a *Global System for Mobile communications* (GSM) network [3]. By making minimal demands to the air interface, the WAP protocol can also operate on low-bandwidth, non-IP GSM bearers such *Short Message Service* (SMS), or *Unstructured Supplementary Service Data* (USSD) channel [4],[5]. Referring to the SMS case, the QoS evaluation for the services provided through WAP is particularly interesting on the basis of the following aspects: (*i*) each SMS message is limited to a maximum of 140 bytes; (*ii*) the logical channel used to convey SMS messages has a slow throughput; (*iii*) WML decks are segmented in several concatenated SMS messages; (*iv*) the transmission of each SMS requires the set-up of a connection between the mobile terminal and the base station. Hence, it is evident the interest for the evaluation of the mean delay to browse a WML deck with GSM-SMS.

This study is an evolution of the techniques studied in the HIPS Project (*Hyper-Interaction within Physical Spaces* of the ESPRIT Program) and it is related to the Information Society Technologies Project, called *Personalised Access to Local Information and services for tOurists* (PALIO), within the 5th Research Framework of the European Commission. This Project is in co-operation with an Italian GSM network operator. PALIO aims at providing multimedia services to tourists for facilitating their visit of a city. At present, tourist information is static and mainly available through paper guides, Internet sites and information points located in specific positions. The PALIO system proposes a wireless access to multimedia services through the mobile phone of the tourist. By means of the WAP protocol a user will be able to receive on a common mobile phone all relevant information stored in a suitable server at the city level. The main experimental services considered by the PALIO system deal with location-dependent personalised information for tourists.

This paper represents a preliminary study related to the mobile browsing of multimedia contents, an interesting perspective for future generation cellular systems.

2. The considered scenario for WAP

We have envisaged a tourist service supported by WAP where WML pages are hosted in a suitable WAP server at the city level. The lowest layer specified by the WAP protocol stack is the *Wireless Datagram Protocol* (WDP) that corresponds to the transport layer of the ISO-OSI reference model. Hence, WAP can be supported by different bearer services in various network infrastructures. Let us refer to the SMS service in GSM networks; we have made such assumption to evaluate the service potentialities in a rapidly deployable configuration that can be supported by all the GSM networks.

The GSM standard envisages various data formats for SMS. The SMS payload is 140 bytes (including, if present, a header of 12-13 bytes) [5],[6]. Even if the SMS channel offers a very low throughput, it must be considered that a WML deck is encoded using a compact binary representation (*tokenised form*) based on the WAP Binary XML (*Extensible Markup Language*) content format [7], thus reducing the deck size on the air interface.

The communication between the mobile user and the WAP server occurs through the *Short Message Service Centre* (SMSC) which is specific for the considered service. The system architecture is shown in Fig. 1, where *ad hoc* SMSC and local WAP server are used in order to reduce the network response time and in order to tailor the WAP server contents to the tourist service managed at the local level. In order to access the services, the user must set on its mobile phone the number of the SMSC.

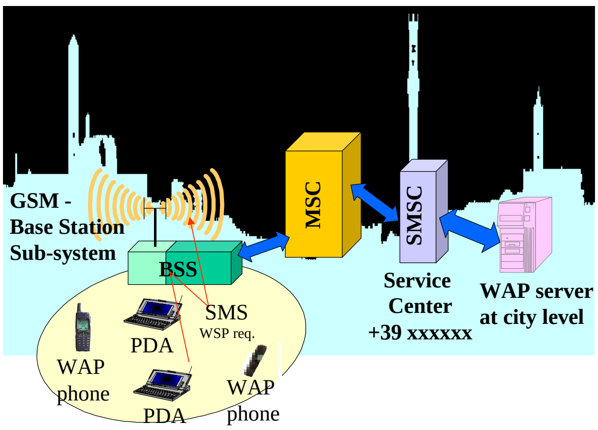

Fig. 1: WAP service scenario.

The user may send its request for a given WML deck with few SMS [2]. The WAP server responses by sending the WML deck in a suitable datagram (the maximum datagram length supported by concatenated SMS is 255 SMS). On the air interface datagrams are compressed according to [7] and are fragmented in SMS sent to the mobile user (here assumed not involved in a phone call) through a *Stand alone Dedicated Control Channel* (SDCCH). SDCCH is a signalling channel that is organised according to a 51-multiframe structure. In particular, in this study we have considered that the first slot of each frame (on only one carrier in the cell) transports the following logic GSM downlink (i.e., from the base station to the mobile users) channels: FCCH+SCH+CCCH+BCCH+SDCCH/4+SACCH/4 [3]. Moreover, we have assumed a configuration where there are only four SDCCH channels, on four consecutive frames in the 51-multiframe. According to the GSM channel coding scheme [8], we have considered that each SDCCH burst conveys 23 information bytes, so that about 6 SDCCH bursts (on a given SDCCH channel) are required to transfer a 140-byte SMS on the air interface.

In this study we have neglected the set-up times for the SDCCH assignment to convey SMS messages in order to focus only on the SDCCH traffic capacity. Hence, for each WML deck requested by the user, the WAP server sends a datagram through a given SDCCH channel as soon as available (we have neglected any other signalling messages to be transported by SDCCH). A *First-Input First-Output* (FIFO) queuing scheme has been assumed for the datagram transmission on SDCCH channels.

3. Traffic model for the mobile browsing with WAP

In order to evaluate the performance of WAP over GSM-SMS, we consider the following downlink traffic model for each browsing user. We assume that the user has started a browsing session and that receives WML decks through a given SDCCH channel. In what follows, we have adapted the models in [9] taking into account the considered scenario. Each user is associated with an on/off downlink traffic generator alternating between an *active period* (i.e., when the user requests decks) and an *idle phase* (that corresponds to a pause in the browsing when the user reads the received decks). The duration of an idle phase is assumed to be exponentially distributed with mean value $1/\mu_{id}$ = 20 s.

During an active period, the user browses a number of decks, which is assumed geometrically distributed with mean value N_d. Referring to our envisaged tourist service based on WAP over GSM-SMS, we have considered cases with 2, 3 and 4 decks per active period (i.e., a short period of browsing). Each deck has been considered as a datagram to be routed to the mobile user. The datagram interarrival time in the active period is exponentially distributed with mean value assumed equal to $1/\mu_a$ = 5 s. The activity factor for this type of source is $\psi_w = (N_d/\mu_a)/(N_d/\mu_a+1/\mu_{id})$ and the mean datagram arrival rate is $\lambda = \mu_a\ \psi_w$ *datagrams/s.* Of course, parameters μ_{id}, μ_a and N_d mainly depend on the behaviour of the browsing user.

We have considered that the mean number of cards per deck depends on both the service type and the bearer service used for WAP. Therefore, we have characterised the distribution of the card length in bytes by measuring the length of more than 1000 cards on different WAP servers. The mean card length resulted as 567 bytes, whereas the mean squared value of a card resulted as 401340 bytes^2. The card length histogram in bytes has been fitted with two distributions: the gamma distribution and the Pareto one [9]. The Gamma distribution has the following probability density function:

$$f_x(x) = \frac{x^{\alpha-1}\, e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \tag{1}$$

According to the fitting process, we have obtained α = 4.4843 and β = 126.6486.

The Pareto distribution has the following probability density function:

$$f_x(x) = \frac{\gamma k^{\gamma}}{x^{\gamma+1}},\, x \geq k \tag{2}$$

Assuming k = 80 bytes (which is about equal to the minimum measured card length), we have fitted the mean value (= 567 bytes) so that we have obtained γ = 1.1. The results of the fitting procedure and the card length histogram are shown in Fig. 2. From this graph, we note that the Pareto distribution does not allow a good fitting, whereas the Gamma distribution appears as a good solution.

We have considered that the number of cards per deck is geometrically distributed with mean value N_c (reasonable values for a prompt delivery of the deck through GSM-SMS are N_c = 3 or 4 cards/deck). According to the content encoding process performed before sending each WML deck on the air interface, we have assumed a mean deck length reduction of 80% (conservative estimate [7]).

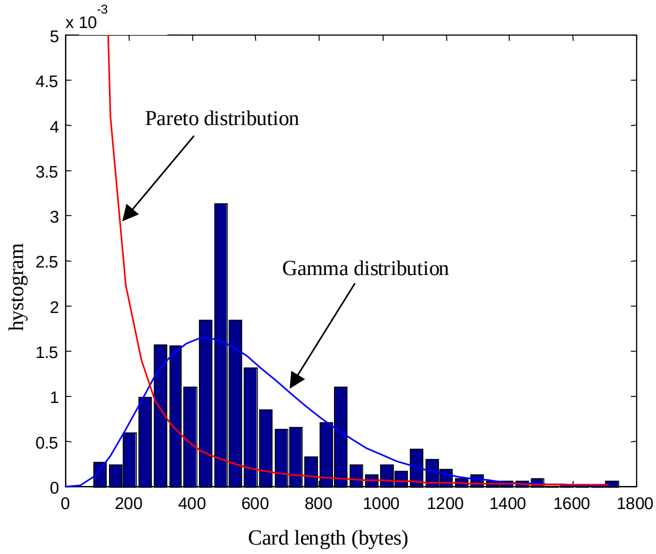

Fig. 2: Histogram and fitting for the card length.

4. Performance Analysis

A downlink performance analysis is presented in this Section to evaluate the mean datagram delay, T_{dtg}, as a function of the number of users involved in the WAP browsing per SDCCH channel, M_w. In this study we consider that a datagram is divided in the equivalent number of "packets", where each packet represents the information content sent on an SDCCH channel (i.e., one slot on four consecutive frames) per multiframe. Hence, a packet conveys 23x4 information bytes.

A datagram traffic source corresponds on downlink to each browsing user. Each source can be approximated by a 2-state *Markov-Modulated Poisson Process* (MMPP), if we make the assumption that the times spent in the active period and an idle phase are exponentially distributed with the mean values presented in Section III. Hence, the modulating process of the resulting traffic source has been shown in Fig. 3; transitions occur at the end of multiframe periods. Accordingly, we have adopted the following system model [10]: $\sum_{M_w} 2-MMPP/D/1$, where: $\sum_{M_w} 2-MMPP$ stands for the aggregation of M_w downlink traffic sources each of them with a 2-state *MMPP* arrival process of datagrams with a length characterised as explained in the previous Section; *D* is the deterministic packet service time; "1" means that only one packet can be transmitted per multiframe.

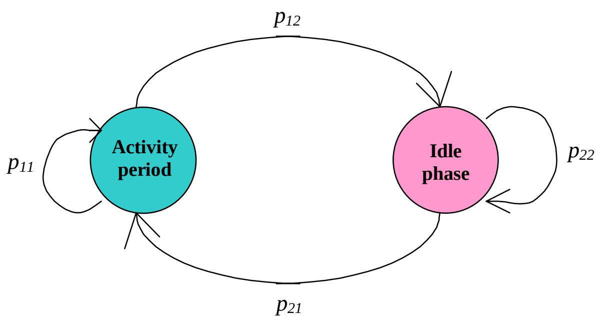

Fig. 3: Traffic source model.

We have embedded the model to the multiframe end instants. Hence, we have modified the approach proposed in [10], by taking into account that packets have a compound arrival process due to both the generation process of datagrams and their variable length in packets. Hence, the packet arrival process is characterised by the following probability-generating matrix:

$$Q(z) = \begin{bmatrix} p_{11} e^{\mu_a T_{MF} [L(z)-1]} & p_{12} \\ p_{21} e^{\mu_a T_{MF} [L(z)-1]} & p_{22} \end{bmatrix} \quad (3)$$

where:

- μ_a is the mean datagram arrival rate during an activity period
- $L(z)$ is the probability-generating function for the length of a datagram in packets (in this case the compressed deck length is considered)
- T_{MF} is the multiframe duration
- $p_{12} = T_{MF}\ \mu_a/N_d$ is the probability that the source leaves the activity phase in T_{MF}
- $p_{11} = 1 - p_{12}$
- $p_{21} = T_{MF}\ \mu_{id}$ is the probability that the source leaves the idle phase in T_{MF}
- $p_{22} = 1 - p_{21}$

According to (3) and [10], we obtain the following expression for T_{dtg}:

$$T_{dtg} = \left\{ 1 + \frac{\rho\left(1 - \frac{1}{M_w}\right) + \frac{\lambda_1''(1) M_w}{\rho}}{2\,[1-\rho]} + \frac{\xi_{M_w}'(1)}{\rho} + \frac{L_d}{2} \right\} T_{MF} \quad [s] \quad (4)$$

where $\rho = \lambda T_{MF} M_w L_d$ erlangs (L_d being the mean length of a datagram in packets) is the traffic intensity produced by M_w users and parameters λ_1''(1) and ξ_{Mw}'(1) have complex expressions that can be found in [10]. For stability reasons the traffic load ρ must be lower than 1.

5. Results

This Section presents the obtained results on the basis of both the traffic model and the analytical approach outlined in the previous Sections. Fig. 4 shows the behaviour of the mean datagram delay, T_{dtg}, as a function of the number of mobile terminals that simultaneously make browsing with WAP per SDCCH channel, M_w. These results have been obtained through formula (4) and by using the statistical characteristics of the length of a datagram in packets according to Section III with $N_c = 3$ or 4 cards/deck (correspondingly, the mean number of packets per datagram, i.e., a deck, is equal to 4.3 , 5.5 and the mean squared value is equal to 29.4 , 50.8). This graph shows three cases for the mean number of datagrams per activity period, i.e., $N_d = 2$, 3 and 4. The mean datagram delay increases with M_w. Of course, the greater N_d and/or N_c the greater the mean datagram delay (e.g., in the case $N_d = 4$, we have T_{dtg} close to 4 s for both $M_w = 5$ with $N_c = 4$ cards/deck and $M_w = 8$ with $N_c = 3$ cards/deck).

It is worth noting that T_{dtg} represents the datagram delay due to the available capacity. An additional delay is also present due to the connection set-up for the transmission of each SMS in which the datagram is segmented. In the considered network configuration with local SMSC, we can roughly estimate that the connection set-up time for each SMS is about equal to two multiframes (one multiframe for the request and another for the response) and that 1 SMS is on average required per card. Anyway, a graph like that presented in Fig. 4 may be useful to show the QoS levels attainable with WAP over GSM-SMS. According to these results, it is possible to design the tourist service envisaged in this study as follows: (*i*) for the GSM operator: dimensioning of the number of GSM signalling channels to provide a service with an adequate QoS level (i.e., T_{dtg} lower than a specified value, for instance 4 s); (*ii*) for the service provider: defining the number of cards per deck so that each user can receive the deck without annoying delays. The results shown in Fig. 4 highlight the limits of the GSM-SMS service as a support for WAP. A further study is required to consider the case where the *General Packet Radio Service* (GPRS) and the related packet switched network elements (GSM phase 2+) are used to convey WAP datagrams.

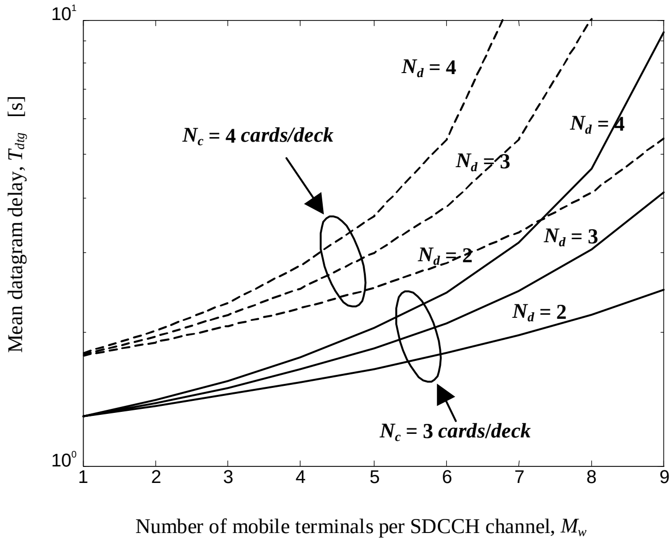

Fig. 4: Mean datagram delay behaviour.

6. Conclusions

This paper has presented a preliminary study on the QoS evaluation for mobile browsing through WAP over GSM-SMS. A tourist service at the city level has been envisaged. Accordingly, a suitable network architecture has been proposed. A model has been defined to characterise the downlink traffic produced by browsing users with WAP. The length in bytes of the cards within decks has been statistically characterised by considering many examples found in different WAP servers. An analytical approach has been described that allows predicting the mean deck delay as a function of the number of users simultaneously involved in WAP sessions. This result permits to estimate the QoS provided to users so that useful considerations can be obtained for designing WAP services.

References

[1] S. Nanda, K. Balachandran, S. Kumar, “Adaptation Techniques in Wireless Packet Data services”, *IEEE Communications Magazine*, pp. 54-64, January 2000.

[2] http://www.wapforum.org/what/WAP_white_pages.pdf

[3] S. M. Redl, M. K. Weber, M. W. Oliphant. An Introduction to GSM. Artech House, MA, 1995

[4] ETSI, “Digital Cellular Telecommunications System (Phase 2+); Point-to-Point (PP) short Message Service (SMS) Support on Mobile Radio Interface” (GSM 04.11).

[5] Wireless Application Protocol Forum Ltd, WAP-158, Wireless Datagram Protocol Specification, Version November 5, 1999.

[6] ETSI - Digital cellular telecommunications system (Phase 2+); Technical realization of the Short Message Service (SMS); Point-to-Point (PP) (GSM 03.40).

[7] Wireless Application Protocol Forum, Ltd, WAP-154, Binary XML Content Format Specification, Version 1.2, 4 November 1999.

[8] ETSI - Digital cellular telecommunications system (Phase 2+); Channel coding, (GSM 05.03 version 5.5.0)

[9] A. E. Brand, A. H. Aghvami, “Multidimensional PRMA with Prioritized Bayesian Broadcast – A MAC Strategy for Multiservice Traffic over UMTS”, *IEEE Trans. Veh. Tech.*, Vol. 47, No. 4, pp. 1148-1161, November 1998.

[10] B. Steyaert, H. Bruneel, G. H. Petit and D. De Vleeschauwer, “A Versatile Queueing Model Applicable in IP Traffic Studies”, *COST 257 Project*, *TD*(00)-02, Barcelona, January 2000.