An Architecture for the Optimum Delivery of Data to a Mobile Client "SMOOTHIE"

Martin Kerrigan[§ϕ], Cormac Sreenan[ϕ] and Fergus O Reilly*

[§]Software Dept.
EMC
Ireland
martin_kerrigan@hotmail.com

[ϕ]Dept. of Computer Science
University College Cork
Ireland
cjs@cs.ucc.ie

*Dept. of Electronic Engineering
Cork Institute of Technology
Ireland
foreilly@cit.ie

Abstract -- **The advent of 2.5G and 3G mobile phone systems will allow a much greater bandwidth capacity for mobile devices than current systems. However, the bandwidth available for mobile systems is constrained by the limited air-interface bandwidth allocated to telecommunications companies operating these systems. As a result, bandwidth over a mobile system's air interface will always be more expensive than fixed line networks.**

Two important issues result. Firstly, operators will need to smoothen out the peaks of demand on the networks. This can be achieved by pre-fetching data at off-peak times that will be requested by a user during a peak period. Similarly, low priority requests that are initiated during peak periods can be delayed until a later time when the system is under a lighter load. Secondly, the operators will need to provide a quality of service assurance to customers who are prepared to pay premium rates for important data transfers.

An architecture of a software system that resides on the client device (SMOOTHIE:- Software Manager for Optimum and Off-peak Transfer through Heterogeneous Interfaces and Environments), to implement mechanisms to address these issues is put forward in this paper. It focuses on requests sent over HTTP and WAP.

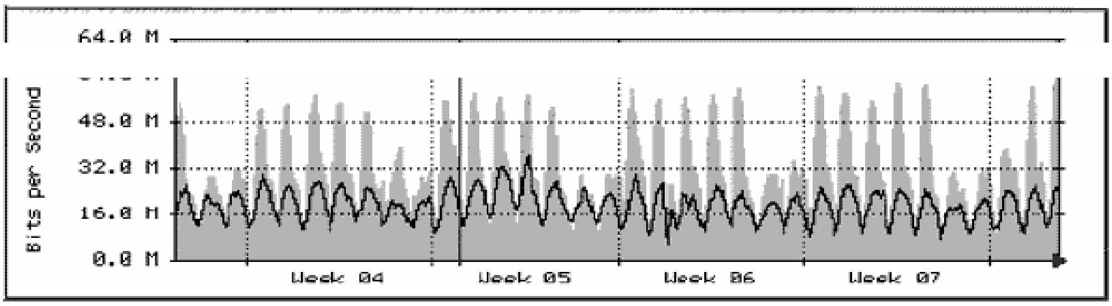
I INTRODUCTION

Much discussion has taken place on what the killer application will be for the next generation of mobile devices. Many now believe that "Access is the Killer App", what matters is not so much what the device can do, but rather the data and computing power to which it has access [1].

If data is supplied only on demand, large peaks in access patterns will mean that there will be time periods when the spectrum for the mobile system is underused, and other times when it is completely saturated. It is important to attempt to exploit bandwidth resources that lie idle at off-peak times. If the transfer of data can be executed at off-peak times and on the cheapest available network, there are potential rewards for both the network provider and the end-user.

Analysis of Wireless Web Traffic [2],[3] has shown that it displays daily peaks in the system. In [2], analysis of the average number of bytes transmitted over a six month period shows a peak demand during the middle of the day, Figure 2.

The cost of a network is generally based on the peak usage rather than the average use. By moving data transfer to the off-peak periods the peak usage would be reduced and hence also cost. The user will be able to avail of off-peak rates and by selecting the cheapest network they will be able to transfer more information for the same budgeted cost.



II OPERATIONAL ASSUMPTIONS

To form the basis for the architecture of a suitable system to deliver information at off peak times, certain justifiable assumptions will be made about wireless networks.

There will always be peak times.
Both fixed and wireless web access patterns studied have shown dramatic increases in access volumes at peak times [2] [3].

The air interface will always be a bottleneck.
Given the difficulty in upgrading the air-interfaces in wireless networks, any system that is aimed at enhancing the capacity of the system must minimise the use of the air interface, particularly at peak times.

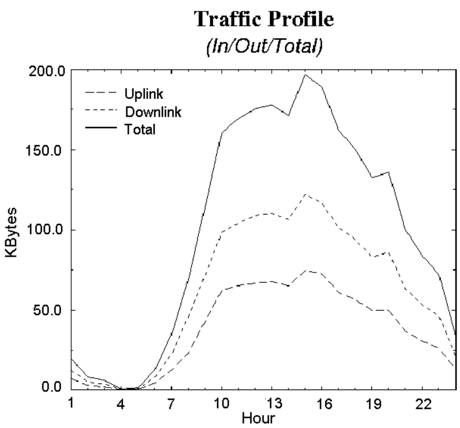
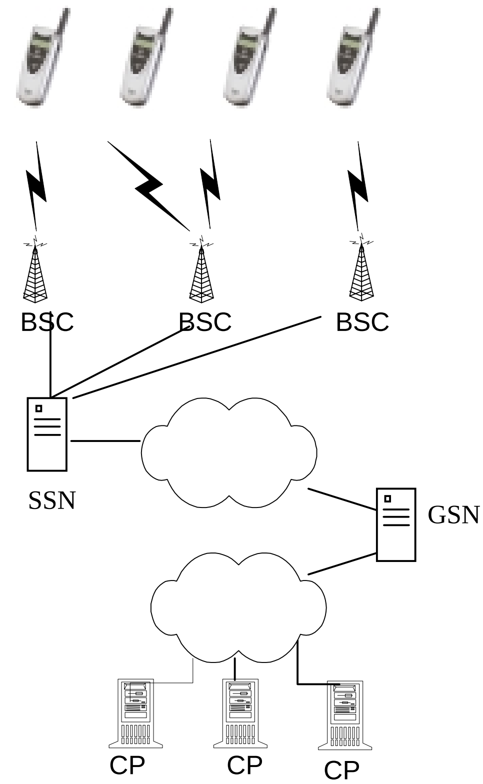


Figure 2: Typical Wireless Web Access Pattern[2]

The air interface is an expensive resource.
As a systems capacity becomes saturated at peak times, operators are likely to introduce new charging schemes based around the peak times of the day.

In the longer term, mobile cellular devices will be able to access multiple networks types.
Of particular significance are the networks operating in the "license free" part of the spectrum used by Wireless LANs and multi-mode terminals [5]. Extensions to Mobile IP to Integrate mobile IP integrate with AAA standards are under consideration for standardization within the IETF. This is discussed in [6]. The BARWAN project [1] investigated how a mobile client may be aware of and use available networks that overlap in coverage.

Data user numbers will increase dramatically. The number of users transferring data over 2.5G and 3G systems will increase dramatically over the next few years. At the end of 2000, there were 39 million mobile Net users in the world, which is set to increase to 729 million by 2005 according to Intermarket Group [7].



BSC: Base Station Controller; SSN Serving Support Node; GSN Gateway Support Node; CP Content Provider

Figure 3: Internet access over the air interface on a wireless network

Many user access patterns will be identifiable.
Web pages that are repeatedly accessed, recently received e-mails, favourites or web resources specifically requested by a user are suitable for pre-fetching. These and the user's personal data will be more predictable than general access to the bulk of information on the Internet [8].

Each user has a unique QoS and cost trade off.
Users will have different maximum additional costs that they are willing to pay for their services. This may also depend on the type of data being transferred.

III HIGH LEVEL REQUIREMENTS OF SMOOTHIE

Periods of disconnection must be allowed for.
Mobile access is often disconnected from the network from time to time. SMOOTHIE must be aware of the patterns of coverage. If the system decides not to download an object to the mobile device until a cheaper time, it must weigh this up with the possibility that coverage will not be available and the object will not be downloadable later. Much of the time, the patterns of coverage will be very repeatable. E.g. an employee may access 'free' bandwidth everyday at the same times in the workplace. However, if the user of a mobile device changes the normal routine e.g. attends a conference, goes to a customer site, a new picture of the pattern of coverage must be built up quickly.

Profiles of a user's interests must be maintained
There are many details about the user that are required. In particular, what resources are likely to be requested by the user, what Internet e-mail providers are used, history of access, bookmarks and topics of interest to the user. Based on history and interests of the user, the system can automatically download information that is of interest to the user at an off-peak time.

Different categories of data must be supported.
Updates to an Internet based personal organisation application require a high priority. E.g a meeting time change requires immediate notification. Some information, for example the news, could be pre-fetched. Other information such as the download of an MP3 file could be delayed. A set of important files may need to be downloaded to a user before a user specified time. Each category of data will need to have an associate priority that can be modified by the user.

User devices must determine the cost of available networks
Since a mobile device will roam between different networks, the cost for data transfer will depend on the available networks as well as between peak/off-peak times. If there is more than one network available at the same time, the cheaper will be the preferred one. SMOOTHIE must make an informed decision as to when is the optimum time to download information. This information also will need to be exposed to the user applications on the client device.

Users must be able to specify a cost model
Cost control will be a different priority for every user. Users will need some method of limiting the cost of using the network. The user must be able to specify what they are willing to pay. A pre-fetching algorithm must be developed to consider this. The user must have the ability to change and configure the cost model.

Minimal changes to Server side
Existing protocols for WAP, HTTP, CIFS and NFS are well established at the server side. Much related work suggested changes to nodes or protocols serving the client device. None of these were widely adopted. Changes to well-established protocols are difficult to introduce as they would require large-scale changes to existing systems or because they would require consensus among a large number of standards bodies and industry players. For a system to gain wide acceptance, it must be transparent to existing protocols and standards.

User Configuration & Intervention
It is very hard to categorise exactly how important different data types are to a particular user. Also, it is hard to specify in a set of rules what exactly is of interest to a user. Even if a complex set of rules are applied, the user may simply change their mind. A user may want to increase the priority of an important data transfer. They may decide that they wish to increase their monthly budget or wish to add a new entry to their list of areas of interest. A user is likely to want to configure or reconfigure the SMOOTHIE system, or intervene on particular data transfers, and the facility to do this must be provided.

Cached data must be stored at the client device
The client device does not have to be the only place data can be cached, however, caching the data at the client device reaps the best rewards.

- One of the most precious resources in a mobile phone system is the bandwidth over the air-interface.
- Because the client device may be disconnected for long periods, important data has to be cached at the client because it may not be accessible when the user requests it.
- While the potential bandwidth available to a client device will increase with 2.5G and 3G technologies, it will not always be available.
    - o The initial bandwidth offerings are not expected to match the indicative minimum

- requirements laid out by The International Telecommunications Union (ITU).
- o It will be shared among users and at peak times and the amount available to a particular user may be small.
- o Some areas may not receive full coverage.
- o Users travelling at high speed will still only receive 144kbs for a 3G system.

IV ARCHITECTURE

SMOOTHIE needs to be able to pre-fetch data and defer requested transfers. When considering web access, the system will sit between the browser on the client device and the air interface. A proxy server on the client device fulfils this requirement and can be based on a well-developed open source web proxy server, e.g. apache [9] and squid [10]. The use of such, well established servers, ensures that all commercially available browsers are supported and tools designed to interface with proxy server can be exploited. The user's browser must be configured to direct requests to this system.

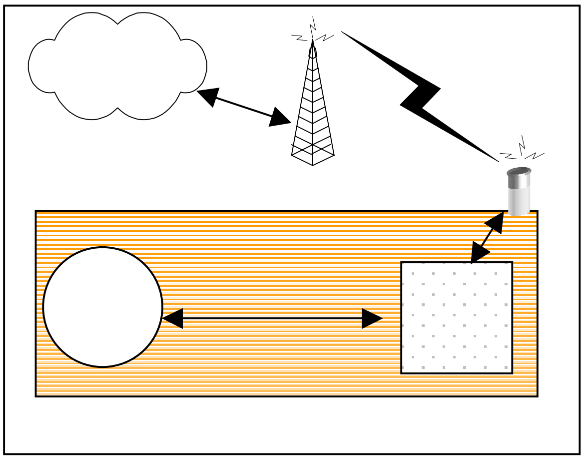


Figure 4 High Level Architecture

Connection Profile Engine/Database

Residing on the client, this will frequently check the network connectivity details and record the information in a database. Cost of particular times would also need to be stored. If the connecting network is not the home network, the cost would depend on two factors:

- Tariff agreements between the home network and available network
- Billing agreement between the user and the home network.

The cost of the data transfer and the probability of connection are important considerations when deciding to transfer information to the mobile device. The Connection Profile Engine will be consulted when making this decision and must make an interface available for this. By making this interface available to other programs running on the mobile device, "SMOOTHIE aware" applications may use it. It may not make sense to initiate activity for an operation if that operation will be cut short due to absence of connectivity to the network or due to the cost of the connectivity. For instance, it may not make sense to begin a networked computer game if it cannot be completed.

Traffic sent over the wire by the connection profile engine is an overhead. The frequency at which the cost of transmission can be checked must be weighed up with the associated overhead. If the network operators can guarantee relatively long intermissions between prices changes, the resources used by checking the cost of transfer can be reduced to a small amount.

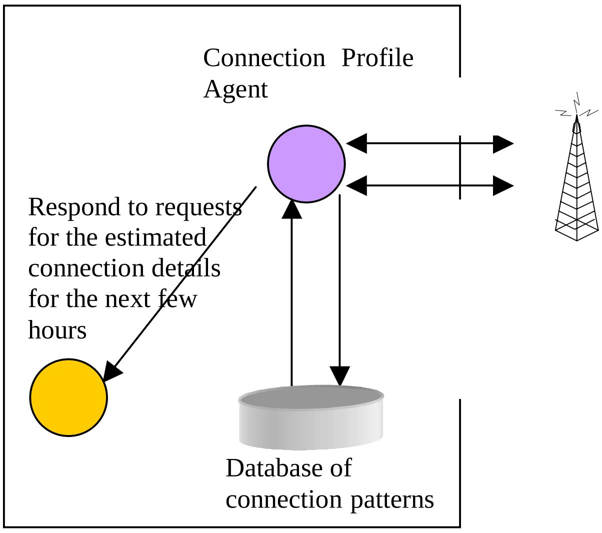


Figure 5 Connection Profile Agent

User Profile Engine

This contains a list of resources that the user may access.

Basic components of a profile include:

- Favourite bookmarked sites
- Historically based sites (last 10, 50 most visited)
- Resources that the user specifies for download.
- An email service specified by the user.
- Specific areas of interest

Sending data of interest to the user is often attempted with simple publish/subscribe models such as that used by MyYahoo. A more advanced profiling system is discussed in [8].

In addition to the basic functionality listed above, a profile engine will need to be able to import other profiles or parts of profiles. This would be used when upgrading to another client device, swapping lists of resources between two sources or restoring the profile in the case of a client device failure.

The profile must also contain information about the user's cost model. This will consist of variables and rules:

- How much are they willing to pay per month.
- What is the maximum rate they are willing to pay.
- What action should be taken when limits are exceeded. Should all transfers be interrupted. Should the current transfer be completed. Should high priority data continue to be transferred.

Pre-Fetch Engine

The pre-fetch engine will have access to the user profile to determine what to pre-fetch. Using a threshold algorithm, an access probability for any particular piece of data is computed. Once this probability exceeds a threshold value, the data is retrieved. Such a system is described in [13]. In addition, the analysis would need to be adapted and developed further so that the probability of access by the user will consider the cost per byte transferred.

Schedule Engine

The schedule engine decides when to send a request to a network based server. It integrates the information from the other systems to implement a decision process based upon

- Cost per byte.
- Connection probability.
- Priority of the data.
- Size of transfer.
- Remaining budget for that billing period.

For pre-fetching another important criteria applies:

- The probability of the data being accessed by the user.

User Interface

In Teleweb [12], a user interface to configure the system is implemented by a process that provides a http interface. HTTP is also used by Marimba [17], a technology that provides management, software distribution and license compliance functionality for personal computing devices.

For SMOOTHIE, a web-based interface is the obvious choice as Internet access is being put forward as the main application to attract a customer base. It is likely that a web-based browser will be available on the mobile device.

V CONCLUSION

High-level requirements and architecture are presented for a system to reduce the peak demand on a wireless system caused by data transfers. What distinguishes it from other previous such systems is

- The probability of connection is considered when deciding to download information to the client device.
- The pre-fetching mechanism includes a consideration for the cost per byte.
- Network operators can dynamically set the cost of data transfers.
- Users can dynamically configure what they are prepared to pay for a data transfers.

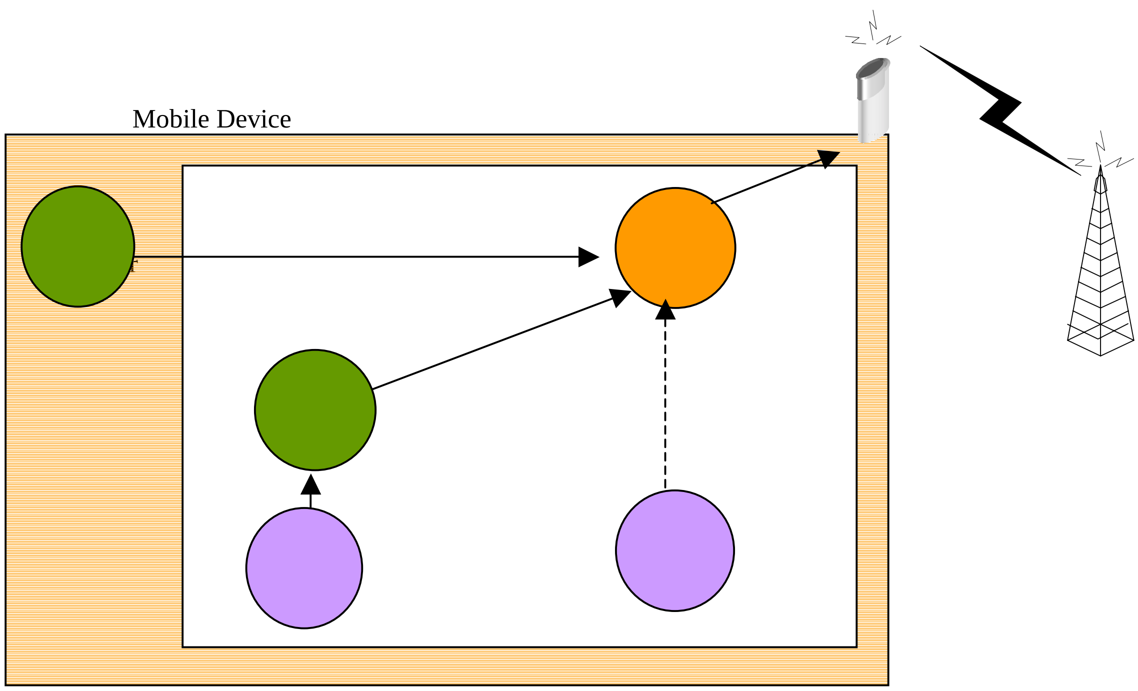


Figure 6: SMOOTHIE

- Users can choose networks based on cost.

This allows for mechanisms to improve cost efficiency of the network for both the network provider and customer:

- Users will be able to avail of off-peak rates and by selecting the cheapest network they will be able to transfer more information for the same budgeted cost.
- The infrastructure cost of a network is generally based on the peak usage rather than the average use. By moving data transfer to the off-peak periods peak usage can be reduced leading to reduced investment costs for operators.

This system can also be used to provide a quality of service for the network. By being able to set the price of the data transfer based on the demands on the system, data that is of low priority can be forced off the system when the system is heavily loaded (the priority here is set by the cost that the user is prepared to pay). This will free up resources for customers who are willing to pay for a high quality of service.

The deployment of the 2.5G and 3G cellular systems presents a financial challenge to the wireless cellular industry [15]. Regulatory bodies vigorously promote competition in this industry. The high cost of the operating licenses in addition to the BTS equipment incurs a heavy financial burden. It may also be necessary to subsidise handsets to build up a customer base quickly. Due to license obligations on levels of coverage, many operators will have to deploy the network with coverage rather than revenue as a top priority. The business case for deploying the next generation system will depend on attracting a large customer base and on getting them to use the new data services. The cost of data transfer for the initial GPRS systems were very high [16] and this may deter further users from adopting the system. By giving the user absolute control of the amount of money they are spending they are likely to feel more comfortable to adopt and use the service. Implementing this at a user application level will give the necessary control and comfort over the amount spent, as well as giving flexibility to change it as often as desired and to a fine level of granularity.

VI REFERENCES

[1] Eric A. Brewer et al, "A Network Architecture for Heterogeneous Mobile Computing", IEEE Per. Commun, Oct 1998.

[2] Thomas Kunz, Thoma Barry, Xinan Zhou, James P. Black, Hugh M. Mahoney, "WAP Traffic: Description and Comparison to WWW Traffic", 3rd ACM International Workshop on Modeling, Analysis & Simulations of Wireless & Mobile Systems, Boston, USA, August 2000.

[3] Atul Adya, Paramvir Bahl, Lili Qiu, "Analyzing the Browse Patterns of Mobile Clients",ACM SIGCOMM Internet Measurement Workshop 2001

[4] http://cc.uoregon.edu/cnews/spring2000/bandwidth.html

[5] M Cristina Ciancetta et al, "Convergence Trends For Fixed and Mobile Services", IEEE Personal Communications, April 1998.

[6] Charles E. Perkins, Nokia Research Center, "Mobile IP Joins Forces with AAA", IEEE Personal Communications August 2000.

[7] http://www.intermarketgroup.com/

[8] Mitch Cherniack, Michael J. Franklin & Stan Zdonik, "Expressing User Profiles for Data Recharging", IEEE Personal Communications August 2001.

[9] http://www.apache.org

[10] http://www.squid.org

[11] http://www.comsoc.org/pci/Public/2001/Dec/index.html

[12] Bill N. Schilit et al, "TeleWeb: Loosely Connected Access to the World Wide Web". Fifth International World Wide Web Conference, 1996

[13] Zhimei Jiang and Leonard Kleinrock, "Web Prefetching in a Mobile Environment", IEEE Personal Communications, October 1998

[14] http://www.cellular-news.com/gprs/tariffs.php

[15] Dimitris Katsianis et. Al, "The Financial Perspective of the Mobile Networks in Europe", IEEE Personal Communications December 2001.

[16] http://www.cellular-news.com/gprs/tariffs.php

[17] http://www.marimba.com/