MAR: A Commuter Router Infrastructure for the Mobile Internet

Pablo Rodriguez
Microsoft Research
7 JJ Thomson Avenue
Cambridge, UK
pablo@microsoft.com

Rajiv Chakravorty
University of Cambridge
Computer Laboratory
Cambridge, UK
rajiv.chakravorty@cl.cam.ac.uk

Julian Chesterfield
University of Cambridge
Computer Laboratory
Cambridge, UK
julian.chesterfield@cl.cam.ac.uk

Ian Pratt
University of Cambridge
Computer Laboratory
Cambridge, UK
ian.pratt@cl.cam.ac.uk

Suman Banerjee
University of
Wisconsin-Madison
Madison, WI, USA
suman@cs.wisc.edu

ABSTRACT

We introduce MAR, a commuter mobile access router infrastructure that exploits wireless diversity (e.g. channel diversity, network diversity, and technology diversity) to provide improved data performance for wireless data users. Our system design stems from the observation that rather than choosing a single wireless service provider (e.g. Sprint, AT&T, BT, Vodafone), a single technology (e.g. GPRS, UMTS, CDMA, 802.11), or a single wireless channel, users can obtain significant benefits by using the multiplicity of choices available. MAR is a wireless multi-homed device that can be placed in moving vehicles (e.g. car, bus, train) to enable high-speed data access. MAR dynamically instantiates new channels based on traffic demand, aggregates the bandwidth and dynamically shifts load from poor quality to better quality channels. MAR, thus, provides a faster, more stable, and reliable communication channel to mobile users.

We have implemented and tested the MAR system in our testbed which spans the networks of three different cellular providers. Through our experiments we have performed a detailed evaluation to quantify the benefits of MAR for different protocols and applications. For example, even in highly mobile environments, MAR, on average, improves the end-user experience of web-browsing and streaming applications by a factor of 2.8 and 4.4 respectively. Our results show that significant benefits can be obtained by exploiting the diversity in coverage offered by many cellular operators, different technology networks (e.g. GPRS, CDMA), and diverse wireless channels.

MobiSYS'04, June 6–9, 2004, Boston, Massachusetts, USA.

Categories and Subject Descriptors

C.2 [**Computer Systems Organization**]: Computer- Communication Networks; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Measurement, Design

Keywords

Multi-link Bandwidth Aggregation, Channel Diversity, Wireless

1. INTRODUCTION

All over the world, 2.5 and 3G wide-area cellular networks (e.g. GPRS, UMTS, CDMA2000) are being deployed to provide high-speed mobile data services. Cellular networks provide wide coverage, enabling ubiquitous wireless data access services. However, they are plagued by problems such as high and variable round trip times, packet losses, burstiness and occasional link outages which significantly limit the final bandwidth offered to the end-user. On the other side of the spectrum of wireless-data technologies, wireless-LAN systems (e.g. 802.11b, bluetooth) are rapidly emerging as an alternative technology to wide-area wireless systems. WLAN systems can easily provide access rates on the order of several Mbps and a much smoother end-user experience than wide-area wireless systems. However, their coverage is limited, mostly restricted to hot-spot areas (e.g. airports, restaurants, hotels, etc.). The end result is that current wide-area and local-area wireless technologies fail to provide the level of bandwidth, coverage, and resilience required to provide high-speed mobile data applications.

Rather than picking a given technology, we argue that a more efficient approach is to use a multitude of wireless technologies simultaneously (e.g. GPRS, UMTS, Digital Video/Audio Broadcast, 802.11). To this extent, we present the design and implementation of MAR - a commuter Mobile Access Router system for *on-the-move* Internet access. MAR is a wireless multi-homed device that can be placed in moving vehicles (e.g. car, bus, train) to enable high-speed data access, thus, creating a *mobile hot-spot*. MAR dynamically instantiates new channels based on traffic demand, aggregates the bandwidth and dynamically shifts load from poor quality to better quality channels. MAR, thus, provides a faster, more stable, and reliable communication channel to mobile users. Local access to the MAR network by users is provided through an 802.11 interface which ensures that the local access link is neither capacity nor wire constrained.

In this paper, we present the method used by MAR in exploiting different levels of wireless diversity (channel, network, technology). Based on thorough experimental tests, we show that significant benefits can be obtained by exploiting the diversity in coverage offered by many cellular operators and different technology networks (e.g. GPRS, 3G). We present the results of a real MAR implementation and quantify the performance of MAR for different protocols and applications in a variety of scenarios. In particular, we show how MAR is able to significantly improve end-user experience for Web browsing and streaming applications in highly mobile environments.

Challenges in Wide-Area Wireless Access

There is a strong growth in mobile Internet access, fuelled by the increasing popularity of *WiFi* (*i.e.* IEEE 802.11b-based WLANs), and the worldwide deployment of wide-area wireless networks such as 2.5G GPRS and third generation wireless (3G). Multi-mode devices (e.g. WLAN-GPRS cards) are becoming increasingly affordable, and a growing number of mobile devices such as laptops, PDAs and handhelds are equipped to connect to multiple networks.

With the proliferation and ever decreasing costs associated with wireless access devices, providers are increasingly looking towards practical issues of service deployment and performance guarantees. Mobility that involves handovers between Wi-Fi 'hotspots', 2.5G and 3G wireless data services continue to pose a significant challenge, as does the intelligent manipulation of channels and multiplexing/striping of data across available wireless links to achieve the best possible performance and access under variable and often unpredictable conditions.

In light of this, it is important for us to identify the challenges in building reliable wireless communication systems:

Wireless Link-related Problems. Cellular networks in the wide-area such as 2.5G General Packet Radio Service (GPRS) and 3G (e.g. UMTS, CDMA2000 etc.) promise users *always-on* connectivity in the wide-area. However, real experiments conducted over production networks (e.g., GPRS, CDMA2000) indicate that such links are currently plagued with several problems such as high and variable round trip times, patterns of burst packet loss, frequent link outages, and significantly lower bandwidths than originally claimed [11],[15]. In other words, it seems that there is currently no wireless technology in the wide-area, that can offer the level of reliability desired. This means that we either have to adapt protocols and applications to work over such links or work for other more practical alternatives.

Spectrum Limitations. Wireless networks are spectrum constrained. Cellular network (and WLAN) operators are only allocated a limited amount of bandwidth. This fixed bandwidth enables them to only support a limited number of subscribers in each service area (cell). However, increasing the data rate for each subscriber is a trade-off against the number of subscribers the service area can support. This situation is particularly exacerbated when subscriber density and application/content size increase at the same time. Licensing laws and competition add to the problem [3].

Lack of Real systems Exploiting Diversity. Most current communication systems are single input single output (SISO) systems; such systems cannot afford to exploit diversity because of the use of only one transmitter and receiver able to operate over the communication channel. Spatial domain solutions can result in considerable improvement in wireless system performance. Techniques that exploit spatial diversity (e.g., Tx/Rx diversity, beam forming, MIMO systems etc.) are proven techniques to improve wireless system (and link) performance. Unfortunately, currently deployed wireless systems have yet to exploit such techniques [3].

After identifying some of the challenges, we feel that it may be difficult (at least in the near term) to realize a wireless communication system that uses a single air interface, and is still able to cater to the requirements of all mobile applications. Instead, we advocate the use of multiple air interfaces simultaneously, to build a better combined wireless communication channel (link). Such a link can provide more predictable data-rates and is better able to meet different mobile application requirements as well as mobility scenarios. To that end, we advocate exploiting network diversity from different wireless networks and operators to be able to aggregate bandwidths that can then be offered as a single large, more stable pipe to end users. We present results from our current implementation of MAR in section 4.

Based on the availability of such wireless diversity, we find that we can make use of MAR to connect to a number of wireless cellular networks simultaneously, and exploit the network diversity in the wide-area to provide a fairly reliable communication link. The advantage from such network diversity is apparent; network interfaces in a MAR router connect to the base-stations of different operators, which are typically sited at different locations or operate at different frequency bands or protocols. In this way, MAR can exploit the inherent network diversity.

2. MAR – ARCHITECTURE

In this section we present the main architecture and components of the MAR system. The MAR system consists primarily of the MAR router which performs bandwidth striping (aggregation) across multiple network wide-area wireless interfaces to exploit the diversity available from different wireless networks to provide a faster, smoother, and more reliable wireless channel. When used in conjunction with a MAR server proxy located in the wired infrastructure (see figure 1) the communication channel can further be optimised to provide transparent TCP, UDP or application based protocol enhancements. Services such as custom-built session striping protocols, more efficient caching and content adaptation and compression might easily be achieved without requiring end system changes.

The core components of the MAR router are presented in Figure 1 and they include a) MAR network adaptation layer, b) MAR Session protocol and c) MAR Proxy Services.

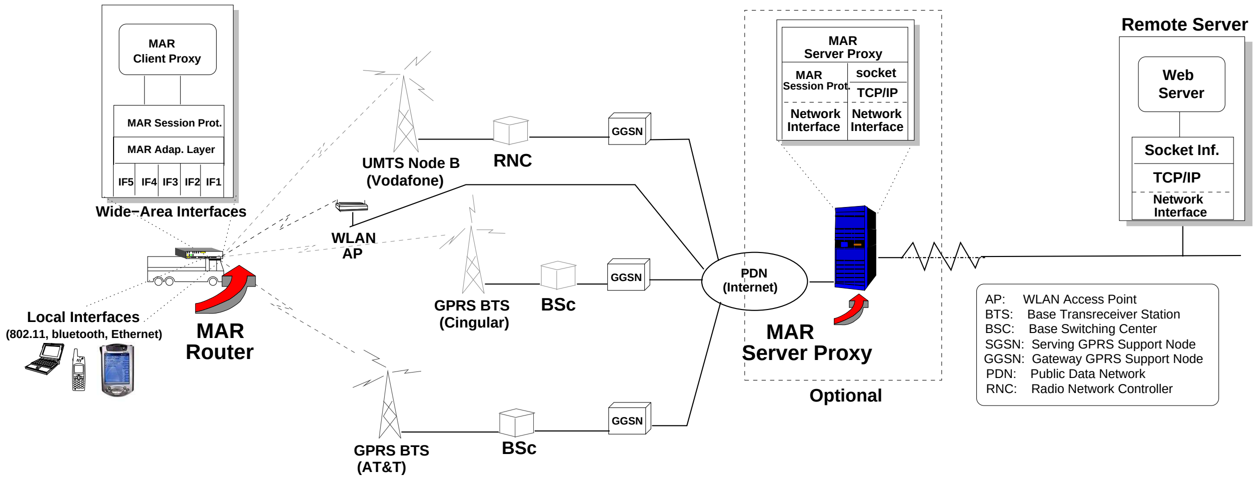

Figure 1: MAR System Architecture and Components.

MAR Network Adaptation Layer: The MAR router provides a set of local interfaces as well as a number of wide-area wireless interfaces. Local interfaces provide access to local mobile users. They can include both wireline as well as wireless technologies (e.g. ethernet, bluetooth, 802.11) to provide connectivity to PDAs, laptops, and other devices. MAR provides a DHCP server that dynamically assigns private IP addresses to local users from a pool of addresses. Mobile devices are configured to use MAR as their default router and DNS server. To this extent, the MAR router runs a local DNS server. MAR also provides a number of wide-area interfaces that can accommodate a variety of wide-area wireless technologies (e.g. GPRS, UMTS, CDMA, long-range 802.11, Digital Video/Audio Broadcast, etc.). The MAR router automatically connects to all preconfigured wide-area networks, authenticates with each operator, and obtains the required parameters to enable each wide-area interface (e.g. IP address, DNS server, default router, etc).

IP addresses assigned by each wide-area operator to each interface belong to different ranges of addresses and in many cases they are obtained from a private range of addresses (e.g. 10.0.0.1, 192.168.2.1). To be able to route requests originating from local mobile clients through the multiple wide-area operators, MAR needs to implement a network adaptation layer. Requests from local users are directed to the MAR router. Based on a particular MAR scheduling policy, MAR selects a given interface for each packet or request. Once a given interface is selected, client-originated packets are source-NATed (both IP address and port) using the IP address of the selected wide-area interface. To the external world, the MAR router appears as a NAT box. Downlink traffic flowing from the wireline towards the clients, will have a destination IP address that is the same address of a given MAR interface. MAR de-NATs such packets and forwards them to the appropriate mobile user.

MAR Session Protocol: One of the key functionalities of MAR is that it can aggregate the available bandwidth across all wireless interfaces. This aggregated bandwidth is then offered as a larger, more stable pipe to the end users. Aggregating bandwidth from the multiple wide-area interfaces is implemented through the MAR Session Protocol.

MAR Session protocol operates between the MAR router and the MAR proxy-server. However, as we will see below, the MAR router can also implement a simple session protocol to work as a stand alone system without requiring the presence of a MAR proxy-server. The scheduling protocol itself is not part of the MAR architecture. Instead, MAR provides an API which can accommodate any custom-purpose built striping protocol (e.g. [14],[16],[17]).

One approach to implement a session scheduler with a MAR router is to use link-layer striping techniques [16, 17], where different packets are sent to each interface regardless of their connection id. However, using such a scheme, different packets belonging to the same TCP connection can end up in different interfaces, and therefore, become different source addresses. Such a scheme will break the TCP connection semantics since servers will receive packets belonging to the same TCP connection with different IP source addresses, hence we see the advantage of using a server proxy to recombine the different channels.

Later in this paper we evaluate a simple MAR Session Protocol that does not require a server proxy. A *per-TCP connection* scheduler is used such that all packets belonging to a given TCP connection are assigned to the same interface for the duration of that connection. In this manner, TCP semantics are maintained, reordering problems are minimized and no changes are required to clients or servers. One drawback with such per-TCP scheduler is that no consideration is given to how much data is requested on each connection. As a result, from the end-users point of view, an end-user may end up waiting for connections that have been assigned large portions of the data, while other connections are idle or being used to serve other users. Solving this requires the use of the enhanced proxy services.

Through the MAR Session API, the MAR router can also accommodate a wide-range of session scheduling policies to load balance packets/connections among different interfaces. Next we describe some of those policies and what elements are exposed by MAR to implement them. The MAR session scheduler can include various policies such as round-robin, least loaded interface, weighted policies, etc. To help determine which interface should be responsible for a given packet/connection, MAR monitors the Signal to Noise Ratio (SNR) and Bit Error Rate (BER) values of each interface and the interface's average throughput. The SNR and BER can be obtained using standard Hayes AT commands available from most wireless cards and phones. MAR uses SNR and BER in addition to throughput information since throughput measurements can be biased by certain protocol behaviours. However, whenever SNR/BER information cannot be collected, MAR relies on past throughput information only. Our experimental results in [26] have shown to report meaningful SNR values, however, BER values can only help to identify severe multi-path interference problems. Note that BER information can only be obtained when data is received.

Based on such information, MAR can determine the weight that should be assigned to each interface to properly perform load balancing (e.g. busier interfaces will be assigned fewer number of TCP connections, while faster interfaces will be assigned higher number of TCP connections). Moreover, MAR can determine whether an interface is in a blackout period or not. For instance if the SNR falls below the minimum sensitivity level or if no packets have been seen for a given period of time, the MAR router can safely assume that a given interface is non-functional. When MAR detects that an interface is going through a blackout period, it stops sending data to that particular interface and re-schedules new packets/sessions to active interfaces. Once a blackout is detected, MAR keeps monitoring the SNR and starts a periodic probe to determine when the interface becomes active again and can be used to aggregate more bandwidth.

MAR Proxy Services: The MAR router can work as a stand alone device or in cooperation with a MAR server proxy located in the wired infrastructure (see figure 1). We envisage multiple uses of a MAR server-proxy including the design of custom-built session striping protocols, TCP optimizations and/or UDP based transport protocol for better link utilization, and application-level optimizations such as more efficient HTTP caching protocols, content compression and adaptation.

Experiments conducted have shown that Internet transport protocols such as TCP are not optimised to provide good end-user experience over wireless wide area networks such as GPRS and 3G [11] [15]. Although the MAR router can work with standard TCP, in order to improve link-layer utilization, the MAR system can also use a MAR proxy-server that acts as a Performance Enhancing Proxy (PEP). Using such a MAR proxy-server, MAR can implement intelligent optimizations that boost the throughput available on each interface. Such optimizations include avoiding TCP 3-way handshake, slow-start, unnecessary DNS queries, spurious time-outs, etc. and are described in [11] [12].

At a higher level, the MAR server-proxy can compress application data objects before forwarding over the wireless links, reducing transfer size and thereby improving response time. Data can be compressed using application-specific lossy as well as lossless compression techniques. Implementations of the MAR system could also include a client side web cache such as squid [27] for leveraging the redundancy in data requests between users over time and thereby reducing the load on the wide-area links and enhancing user performance. We present these options as future extensions to the system without detailed analysis in this paper.

3. EXPLOITING NETWORK DIVERSITY

The ability to provide sustainable connectivity and data rates using multiple wireless access technologies available in the same mobile terminal is a challenge. This is made significantly more difficult because of the wide variety of environments (indoor, outdoor, moving, fixed, etc.). In this section, we discuss our practical findings of *how* and *how much* network diversity can be exploited in wide-area wireless environments.

GPRS and 3G links like all other wide-area wireless networks, exhibit many of the following characteristics: low and fluctuating bandwidth, high and variable latency, and occasional link 'blackouts' [15][9]. Previous work has given insight into the characteristics of these links. However, we are aware of no publicly available work that has quantified the extent of the diversity in the wide-area wireless environments. In this section, we conduct a series of wireless diversity tests. These have been repeated under a wide range of conditions – from stationary (in-building environments) to highly mobile (bus, train etc.) using different models and manufacturer of handsets, and different sets of network operators. These tests corroborate our finding on differences in terms of coverage being offered by the operators, harsh conditions of the mobile environments, and also its impact on application performance. In this section we summarize our findings. A comprehensive description on diversity tests is available in the form of a separate technical report [26]. We examine three different classes of diversity that are available, which we call technology diversity, network diversity and channel diversity respectively:

- **Technology Diversity:** In technology diversity we consider the (dis)-similarity of the data performance observed by two wireless devices that are operating using two different communication technologies, e.g. CDMA and GPRS. GPRS is a data bearer service for the GSM system that uses a Time Division Multiple Access scheme to assign separate channels to different mobile devices. We expect that in urban areas, a MAR device with two interfaces that use different communication technologies, would experience minimal correlation in the data performance between the two interfaces.

- **Network Diversity:** In network diversity, we examine the data performance of different wireless devices that are using the same underlying technology, but are attached to the networks of different operators. A MAR device with multiple interfaces that are attached to the networks of different operators typically observes uncorrelated data performance on these interfaces. This is primarily because each interface connects to a different Base Station, and the wireless channel properties of the path between the Base Station and wireless interface are independent of each other. Even where operators share the same base station location the frequency band or orientation of the antenna will likely be different.

- **Channel Diversity:** Finally, in channel diversity we examine the differences in wireless data performance between multiple interfaces that are connected to the network of the same wireless cellular operator. A Base Station of a cellular network simultaneously assigns different channels to multiple devices. The channel assignment technique is different in different cellular systems, e.g. different time-slots in GSM-based systems, different spreading codes in CDMA-based systems, etc. In many cases two interfaces of a MAR device that are attached to the same cellular network, will connect to the same Base Station. Therefore the correlation in data performance observed by these two interfaces will be relatively higher than the technology or network diversity cases.

3.1 Quantifying Diversity

We conducted experiments to quantify the available diversity of two wireless devices along these dimensions for both static and mobile scenarios.

Setup

Figure 2 shows the experimental setup for our tests. We used a MAR router with multiple interfaces. For these experiments we used four different operators that provide wide-area wireless data services in our area (Opt1, Opt2, Opt3, and Opt4). Three of these operators (Opt1, Opt2, and Opt3) provide GPRS data services on the 1.8-1.9 GHz (Opt2) and 800-950 MHz band (Opt1, Opt3). The fourth (Opt4) operator provides CDMA 1xRTT services on the 800 MHz band. For different experiments we used a variety of hardware to simultaneously connect multiple interfaces to the same or different operator's networks. In particular, we used three PCMCIA Sierra Wireless cards with '4+1' slots (4 downlink GSM slots, and 1 uplink), and two Motorola Phones (Motorola T260 Phone, '3+1' handset) that were connected through serial-line interfaces to the router. For the channel and network diversity experiment, we performed experiments using the live-production networks of the three GPRS operators. In the technology diversity experiments we used GPRS and CDMA 1xRTT as the two underlying communication technologies.

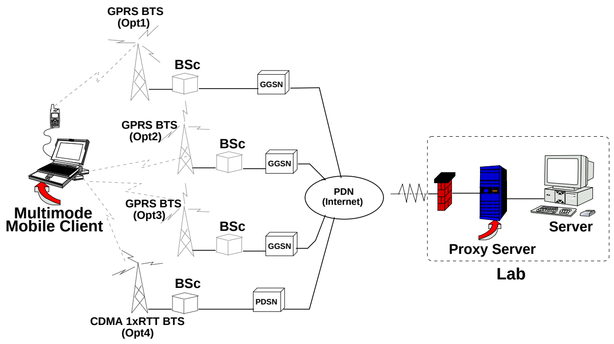

Figure 2: Experimental Set-up for Diversity Experiments with TCP.

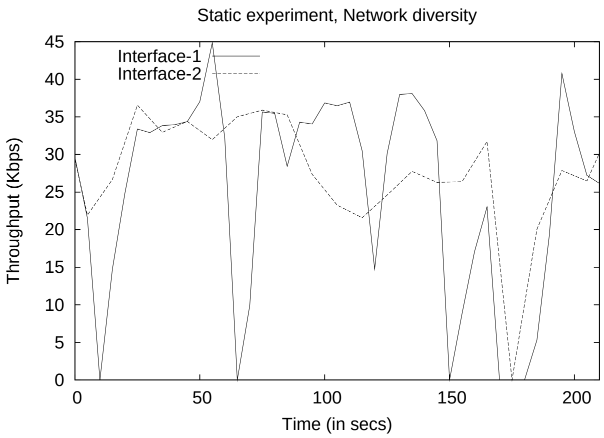

Figure 3: Instantaneous throughput variation for two co-located wireless devices connected to different operator's network (Opt1, Opt2) (an example snapshot from a static experiment).

To quantify the diversity existing among different channels, networks, and technologies, we simultaneously initiated an `ftp` download for a large data file through each of the interfaces. The file was being hosted in a server in our lab and all traffic was going through a proxy-server. This proxy-server implements a number of TCP optimizations to improve TCP link utilization over the wireless links. These optimizations included avoiding slow start, removing spurious time-outs and decreasing traffic burstiness. A complete description of the optimizations implemented in this proxy can be found in [10].

The static scenarios were performed at multiple locations in the city (including our laboratory). The mobile experiments were performed in a car moving through the city streets.

The duration of each experiment was approximately 30 minutes. We collected `tcpdump` logs at the MAR client to monitor the progress of the downloads. Subsequently we used these traces to calculate the instantaneous throughput (computed as the average data transfer rate for each sequence of 10 TCP segments).

Results

For a static experiment with two wireless cards connected to different operators' networks, we show a snapshot of the instantaneous throughput of the two devices in Figure 3. (The figure shows a part of the entire download.) This is an example of network diversity. Both the cards show occasional loss of throughput. However these occurrences are independent of each other as can be observed in the plot (Interface-1 experiences decreased throughput around time instants 10, 65, 150 and 168 seconds, while Interface-2 experiences the same around time instant 172 seconds).

Figure 4 shows a snapshot from another experiment of network diversity for a mobile scenario. In this example, Interface-2 experiences a large throughput outage period between time instants 40 to 220 seconds, while the throughput of Interface-1 remains unaffected.

In Figure 5 we show a snapshot from a static experiment for the channel diversity case, i.e. both cards connect to the same operator's network. We can see that loss of throughput is more correlated in this case than the network diversity experiments. For example, both the interfaces experience a loss of throughput at time instants 20, 125, and 180 seconds. However, there are other instants of time where each

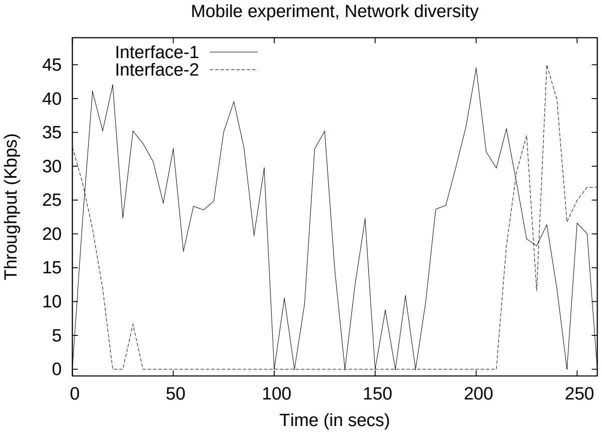

Figure 4: Instantaneous throughput variation for two co-located wireless devices connected to different operator's networks (Opt1, Opt2) (an example snapshot from a mobile experiment).

interface independently experiences loss of throughput.

To measure the overall correlation in different such experiments, we use the Pearson correlation coefficient. The Pearsons's correlation coefficient (ρ) expresses the degree of linear relationship across all of our data between two variables, X and Y and is given by

$$\rho = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

where N is the number of data points. ρ varies between -1 and +1, where -1 is perfectly negative correlation and +1 is perfectly positive correlation. A value of 0 indicates no correlation.

Scenario	Pearson's coefficient (ρ)
Mobile, Channel diversity	0.28
Static, Channel diversity	0.14
Mobile, Network diversity	-0.18
Static, Network diversity	-0.04
Mobile, Technology diversity	-0.03
Static, Technology diversity	-0.02

Table 1: Pearson's correlation coefficient to quantify diversity in various scenarios. The scenarios are sorted in decreasing order of ρ.

We tabulate the values of ρ for different experiments in Table 1. All the correlation values are either moderate, low or negative. This implies that there is significant potential for MAR to exploit different kinds of diversity in its operations. We can observe that the channel diversity case has the highest correlation under mobile conditions as would be expected, while technology and network diversity consistently exhibit correlation values of less than or equal to 0.2 indicating little or no correlation. For more a complete diversity study, including a larger set of operators, longer traces, and parallel SNR/BER measurements, please refer to [26].

4. MAR PERFORMANCE

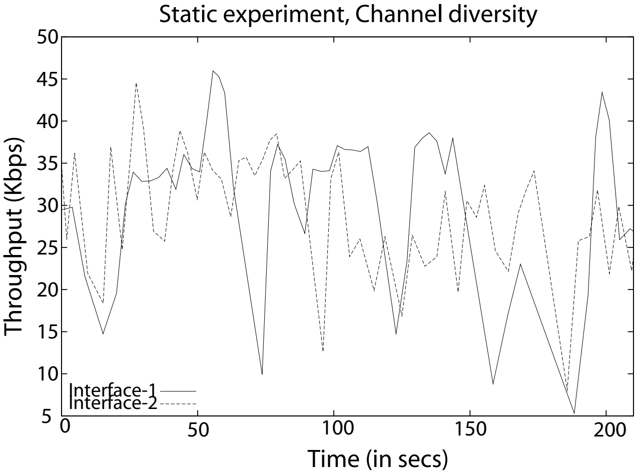

Figure 5: Instantaneous throughput variation for two co-located wireless devices connected to the same operator's network (Opt1) (an example snapshot from a static experiment).

In this section we study how the MAR system uses diversity to improve link throughput and robustness. We consider both mobile and static environments. In Section 5 we then study how these results affect the performance of several popular applications using a MAR system.

4.1 Implementation and Experimental Testbed

We implemented the basic MAR router functionality on a Linux based platform (we also implemented similar functionality as a Windows driver for Windows XP). This implementation provides the MAR network layer adaptation, uses TCP as transport protocol, implements the basic MAR session protocol that uses per-TCP connection state scheduling, and provides a local caching system (Squid). Our MAR router is built with off-the-shelf components and is equipped with multiple wireless interfaces to accommodate several wide-area network carriers. In Figure 6 we show the basic setup for our experimental data collection, which is similar to the one described in Section 3.1.

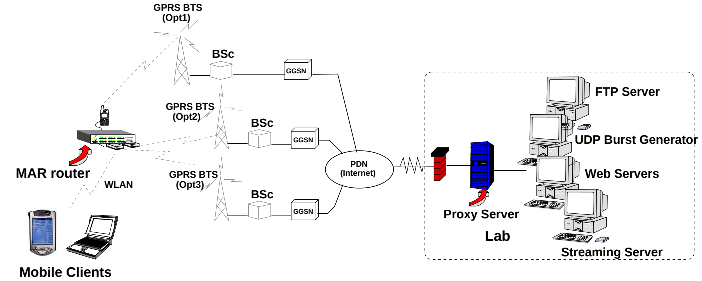

Figure 6: Experimental Set-up (for all applications).

For mobile clients, we used a number of mobile devices including laptops and PDAs. Client devices did not require any manual configuration and they were configured to use the MAR router as their default router and DNS server via DHCP. The local Squid cache running on the MAR router was working in transparent mode, therefore, mobile users did not have to manually point their browsers to benefit from the MAR caching system.

For the following experiments we used the three cellular wireless operators that provide GPRS services in our area (Opt1, Opt2, Opt3). For the mobile tests, we placed the MAR router on a car and drove in urban areas. The average approximate speed of the car during the experiments was 20 mph.

4.2 Throughput

In this section we characterise the MAR link throughput in both static and mobile environments, determining the impact of diversity on the overall MAR throughput. To this extent, we use an open-loop UDP stream rather than TCP to ensure that our measurements were not biased by any particular TCP-wireless behaviour. To evaluate the performance of UDP traffic in the MAR system, we created a client-server application that measures the available UDP throughput by sending packets at a rate close to the link capacity. This network application sends bursts of data packets back-to-back from the wire-network towards the wireless Base Stations and measures throughput based on the inter-arrival times between packets in a burst. Back-to-back packets are buffered by Base Stations and forwarded to the mobile device at the wireless link rate until the buffer is depleted.

Given that we were using three different GPRS operators, we opened three parallel UDP sessions between the UDP server and the mobile client. Each UDP session used a different interface. Packets were pushed to the mobile device on all sessions. The MAR system was able to receive packets from all three interfaces simultaneously.

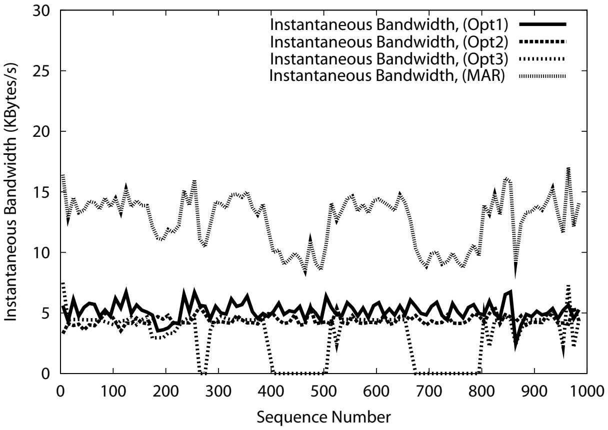

Figure 7: MAR UDP Performance. Static environment. Throughput for each interface stand alone, and for MAR

In Figure 7 we show the UDP throughput for three different GPRS cellular operators. Such throughput was calculated over 200 bursts of 5 packets. The results show that the UDP throughput of two interfaces was quite stable around 5 Kbytes/sec, however, the third interface was experiencing frequent blackouts, where no data was received. In this Figure, we can also see the performance of the MAR system. We observe that the MAR throughput is much higher than the throughput of the best interface, around 14 KBytes/sec.

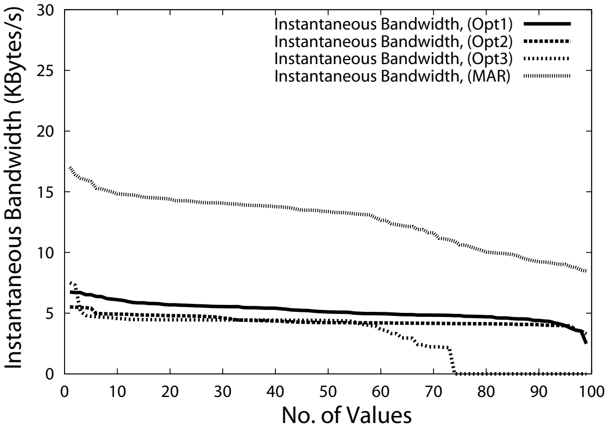

Figure 8: MAR UDP Performance. Static environment. Aggregated sorted throughput.

We can also see that diversity ensures that the MAR system does not see any blackout periods, even during the periods when the third interface is not working well.

To better understand the throughput provided by different operators, in Figure 8 we present the same values obtained in Figure 7 ordered by throughput from highest to lowest. From this Figure we can see that interfaces 1 and 2 provide a very stable throughput for most of the trace. Only at the end of the Figure, we can see that these two interfaces had a number of samples with low throughputs or no data. On the contrary, interface 3 has a rapid decay in its throughput, thus, having many samples where throughput was zero. MAR has a very stable throughput around 14 KBytes/sec and does not go below values lower than 11 Kbytes/sec. From this figure we can see that at a given location, picking the wrong interface may provide very bad end-user experience. However, MAR can easily provide a throughput that is much better than the best throughput provided by any interface.

In Figure 9 we repeat the same experiment as in Figure 7 in a mobile environment to better understand how MAR benefits from diversity in a more hostile scenario. From this Figure we can see that the throughput variability is much higher than in a static environment, with sudden changes in the operator's ranking. Average throughputs for interfaces 1 and 2 remain at approximately 5 KBytes/sec, while interface 3 still provides a significantly lower throughput than the others. Given the high variability imposed by the mobility environment, a client connected to a single interface will experience rapid changes in throughput and suffer from frequent blackouts. For a more intelligent client that automatically selects the fastest interface, selecting the best interface becomes much more complicated since it is harder to identify a provider that consistently provides better throughput than the rest. Even for a client that could automatically select the fastest interface at any point in time, it would not be able to enjoy more than 5 Kbytes/sec on average. On the other hand, with MAR, the throughput is much higher than with any other interface alone, providing an average throughput of around 10 Kbytes/sec without requiring any sophisticated channel selection mechanism. In addition, throughput is more stable, efficiently preventing low throughput areas and blackouts.

Figure 8 reproduces the results presented in Figure 9 sorting the instantaneous throughput from highest to lowest in a mobile environment. We can see the sorted throughput for all interfaces in a mobile environment drops faster than in a static environment (Figure 8). Thus, the amount of time spent in the lower throughput areas is higher for each individual interface. On the other hand, for the MAR system throughput stays high for most of the time and it never falls to zero. Even in such a mobile environment, the impact of highly variable rates in the MAR system is not as pronounced as for each individual interface due to the fact that MAR exploits the benefits offered by different types of diversity.

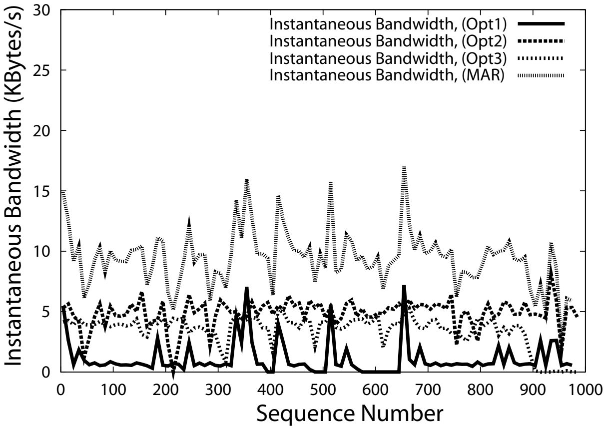

Figure 9: MAR UDP Performance. Mobile environment. Throughput for each interface stand alone, and for MAR

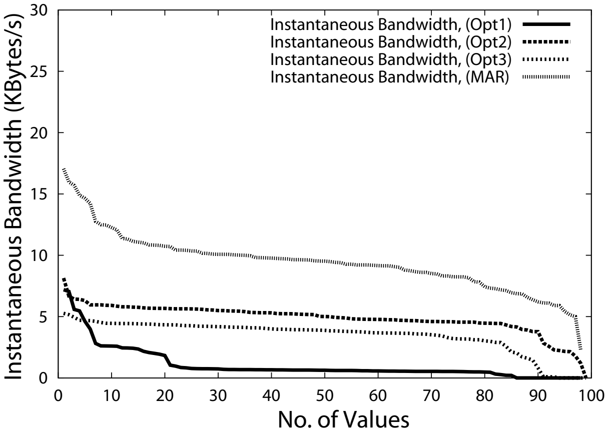

Figure 10: MAR UDP Performance. Mobile environment. Aggregated sorted throughput.

4.3 MAR Blackout Reduction

One of the main factors that affects end-user experience is the frequent presence of blackout periods in most operators [10]. These blackout periods are normally due to interference problems, hardware failures, or loss of connectivity or coverage. In this section we quantify the performance of MAR under blackouts. To this extent, we define a blackout period to be a period of time greater than a given *blackout threshold* where no data is received. To measure the blackout periods we collected traces for UDP traffic in static and mobile environments during a period of 1800 seconds for each interface and for the MAR system. For each of these traces we identified those periods of time where the interarrival time between packets was greater than a given blackout threshold. In the next table we show the percentage of time spent in a blackout period, assuming a blackout threshold of 10 seconds.

Table 2: Amount of time spent in a blackout by each operator and by the MAR system. Mobile environment. Blackout threshold equal to 10 sec.

Protocol	Opt1	Opt2	Opt3	MAR
UDP (Static)	2.7%	1.3%	3%	0%
UDP (Mobile)	6%	3%	4%	1%

From this table we can see that the amount of time spent in a black out can be quite significant for a given operator (e.g. between 3% to 13% of the total trace time for Opt3). Mobile environments also experience on average 2.6 blackout periods more than their counterpart static environments. This is a natural effect of the fact that under mobile environments, mobile devices suffer frequent cell handoffs, loss of coverage, and sudden disconnections.

When comparing the above results with the percentage of time spent by the MAR system in a blackout, we see that the MAR router spends a much smaller portion of the time in a blackout. Thus, the probability that the MAR system cannot receive data from any of its interfaces for a period of 10 seconds or more is almost negligible. Only in mobile environments, MAR experiences small blackout periods that account for 1-2% of the total trace time. Therefore the MAR system significantly increases resilient against network failures. This probability should be smaller and smaller as the number of interfaces increases. This improved resilience behaviour of the MAR system is due to the high diversity shown among multiple operators, channels, and technologies presented in Section 3.

5. MAR APPLICATION PERFORMANCE

In this section we study the performance of MAR for two important applications to better understand the impact of MAR in the actual end-user experience. The two applications considered are Web downloads and Video/Audio streaming.

5.1 Web Performance

To study the performance of Web traffic we considered three popular Web sites (CNN, Amazon, and Yahoo). For each of these Web sites we replicated their front page content in our lab. This was done to avoid the effects of content updates, along with the performance vagaries of the public internet. We downloaded and hosted all necessary objects on our own web servers, creating multiple virtual hosts as necessary. We also hosted our own DNS server with all necessary records to reproduce the exact setup of the target Web pages. Experiments were performed in a mobile scenario. Each page was downloaded 20 times and the total page download time was averaged over all samples. To this extent, we instrumented the Internet Explorer 6.0 browser to calculate the total page download time and log it into a file. The maximum number of connections allowed by the browser was 30 and both browser and server supported HTTP/1.1 persistent connections. Connections opened by the browser were scheduled using a per-interface round-robin policy. All packets from the same TCP connection were assigned to the same interface. No explicit proxy was configured at the browser. The characteristic of each page are: CNN (172 KBytes, 88 objects, 6 different embedded domain names), Amazon (94 KBytes, 43 objects, 3 different embedded domain names), Yahoo (61 KBytes, 17 objects, 3 different embedded domain names).

To understand the impact of MAR in Web browsing, we performed the following measurements: We first measured the download times seen by the mobile user using one single interface for different operators (Opt1, Opt2, and Opt3). Then, we downloaded the same pages using the MAR router with three parallel interfaces and the same set of operators. Finally, we calculated the optimal MAR response time. The MAR optimal response time is calculated by summing all throughputs from all parallel-interfaces during a given download. This optimal response time corresponds to that provided by an optimal session-level scheduler.

In Figure 12 we show the average response time seen by the mobile browser for different individual interfaces, for the MAR system with the three above interfaces, and the optimal response time. From this Figure we can clearly see that the response time provided by each individual interface is always much higher than for the MAR system. On average, users browsing Web content behind a MAR router can experience an acceleration factor of 2.8. This is a very significant improvement for mobile users. For instance, a page like Yahoo that can take to download almost two minutes without MAR, can be retrieved in less than 40 sec with MAR. If the number of MAR interfaces is increased, the response times can be decreased even further.

In Figure 12 we also quantify the response time when MAR uses a client proxy-cache. When MAR caching is turned on, requests from the mobile browsers are transparently redirected to the cache. Requests that result in a cache-HIT are satisfied very fast through the local-area interface, while missing objects are retrieved from the origin server. To understand the improvement provided by the MAR caching system we repeat the experiments in Figure 12 using MAR client-proxy cache. The results show that caching can clearly have a significant impact in reducing end-user latency, improving the average response time provided by MAR by an additional factor of 1.8. Even for a page like CNN that is known to have a lot of uncacheable content, MAR client proxy-caching still provides significant benefits since cached portions are delivered very fast.

Finally, in Figure 12 we compare the response time offered by the sample per TCP-connection scheduler implemented in MAR with the response time obtained from an optimal scheduler. From this Figure we can see that this simple scheduler performs quite well and provides a response time that is not far from the optimal one. However, in some cases this scheduler showed important weaknesses that need to be overcome by more efficient session scheduling protocols.

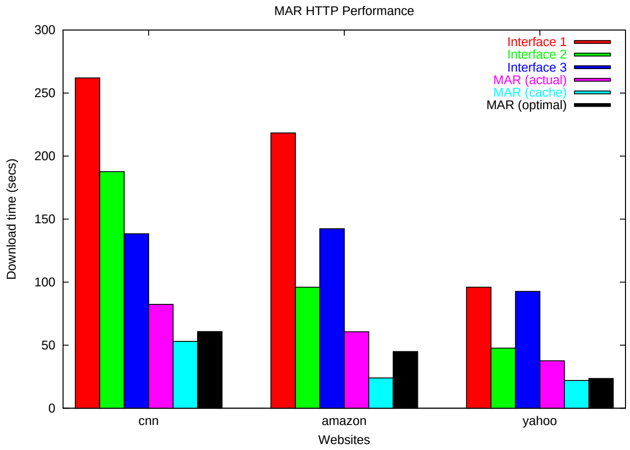

Figure 12: MAR HTTP Performance. Latency for different Web sites where the first three columns in each set represents the performance of each operator stand alone, next one is MAR (current implementation), next one is MAR with caching, and the last one is MAR (ideal).

For instance, in Figure 11 we can see the HTTP timeline distribution for each of the three interfaces used by MAR to download the CNN Web page. Each HTTP timeline plot provides the evolution of each TCP connection opened by the browser through a given wide-area interface. From this Figure we can see that the number of TCP connections opened by the MAR scheduler through each interface is about the same and equal to 11 connections. Thus, the MAR scheduler is doing a good job of load balancing requests among interfaces. However, we can see that not all interfaces finished at the same time. While the first two interfaces finished their download around 40-50 sec, the third interface kept downloading data all the way up to 70 sec. As a result, the end-user had to wait for the slowest interface to finish and could not benefit from the other idle interfaces.

This is due to the fact that many large objects were assigned to the third interface, rather than being spread out over all interfaces. This is an interesting observation which clearly shows the limitations of such a simple scheduling protocol. Finding a session-level protocol that provides an optimal performance for the MAR router requires a thorough analysis of session-level schedulers under different deployment scenarios and applications. However, such a discussion is beyond the scope of this paper and the subject of on-going work.

5.2 Streaming Performance

In addition to improving HTTP throughput, we consider the potential for a MAR system to also improve streaming performance. To quantify the improvement provided by MAR under a streaming scenario we implemented a UDP streaming system that sends packets through multiple interfaces in parallel, and evaluated the streaming quality for each interface and for the MAR system in terms of throughput, jitter, and buffer starvation.

To this extent, we consider a MAR streaming server that efficiently utilises the available bandwidth from all channels,

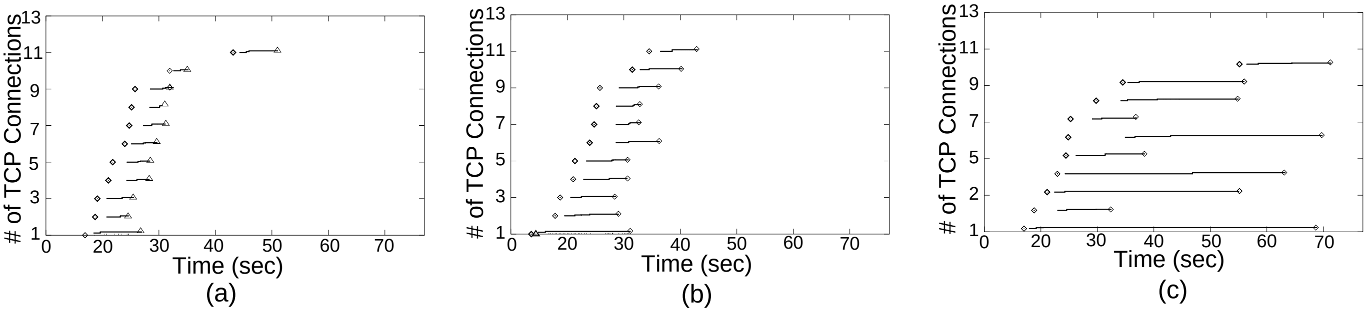

Figure 11: HTTP time line sequence for each wireless interface while downloading CNN top-level page. a) Opt1, b) Opt2, c) Opt3. Each small rise in the lines represent a separate GET request made using that specific TCP connection.

ensuring that the streaming application closely matches the capacity of the links. MAR streaming server uses a rate allocation and a partition algorithm that minimizes the probability of packets arriving late. For each interface, the server estimates the average interface rate using the standard moving weighted average. Based on the estimated rate, the streaming server determines which packets should be sent on which interface. Thus, to minimize the chances of packets arriving late, slow interfaces are assigned packets to be played further in the future, while fast interfaces are assigned packets to be played in the near future. This assumes fairly accurate flow control, i.e. packets are only sent on links which have available capacity thereby avoiding unnecessary buffering during blackouts and packets reach the receiver in time.

When the rate estimation algorithm fails and a certain interface becomes much slower than predicted, packets are lost. To relax the dependency on exact channel rate estimation, we allow for a certain initial buffer. To offset the effects of the variable bandwidth over individual links we also apply FEC techniques over the data. Such an approach introduces a fixed amount of overhead on the data being transmitted, but presents the desirable property that for an (n,k) code, by encoding a data object into n blocks of data, only k blocks out of any n are required to reconstruct the original source data. Consequently, for a combined channel that statistically can support a certain streaming rate, as long as the encoded data size is less than the expected capacity of the channel over a period of time equivalent to the data sample period, we would expect to be able to reconstruct the source data. More details on how to implement a streaming server that efficiently makes use of multiple interfaces can be found in [24] [25].

On this basis we consider the potential performance of such an algorithm using real UDP traces collected on a mobile environment. We therefore quantify the streaming performance seen by an end-user for a single interface under multiple GPRS providers, and for the MAR system using three parallel GPRS interfaces (Opt1, Opt2, and Opt3).

The streaming session duration was approximately half an hour, which corresponds to a mid-size clip. We simulated a streaming application where mobile users target a video streaming rate that is determined by the average throughput available through their connection. Thus, users on a single slow interface (e.g. 20 kbps) target low quality videos with average throughput equal to the throughput of the interface (i.e. 20 kbps), while mobile users using MAR target much higher rates. We assume that before starting playing a given video, an initial play-out buffer of 10 seconds is built for all videos.

We assume 25 Kbyte data chunks where we will perform FEC encoding. We consider a value of (n-k) to be equal to 2, which will allow us to recover for almost all packets lost in our trace. The size of each individual block in this example would likely be something close to the MTU of the link, so for a 1500 byte MTU, we could achieve payload sizes of 1470 Bytes which would generate 17 blocks maximum, or in our example a (17,15) code. Our useable throughput is therefore going to be a factor of $(n-k)*B$ where B is the block size, e.g. in this example 88.2%.

Given these parameters, in Figure 13 we plot the cumulative streaming data received by each interface and by the MAR system at any given time during the duration of the session. We clearly see that using MAR, the amount of useful streaming data received by a given time is much higher than the amount of data received using any of the other stand alone interfaces. Thus, a mobile client can target much higher consumption rates using MAR than without it.

Not only MAR enables faster streaming rates, it also provides a much smoother streaming experience since it decreases the burstiness of the stream received by the end user. To show this effect, in Figure 14 we plotted the inter-arrival time distribution between video frames for a single interface and for MAR. Inter-arrival time distributions for Opt1, Opt2, and Opt3 are very similar with an average inter-arrival time around 450 msec and a long tail. For MAR, we can see that a) it provides a much smaller average inter-arrival times around 154 msec, which enables higher streaming rates, and b) it provides a lower inter-arrival time variance, thus, decreasing the burstiness seen by the streaming player. This benefits should be more pronounced as the number of interfaces multiplexed by the MAR system increases.

To better quantify these effects in the end-user streaming experience, in Figure 15 we show the amount of outstanding data buffered in the streaming player for each individual interface and for the MAR system. Whenever, the amount of buffered data falls below zero there is buffer starvation and the video playout will freeze until the buffer size becomes greater than zero. From this Figure we see that using

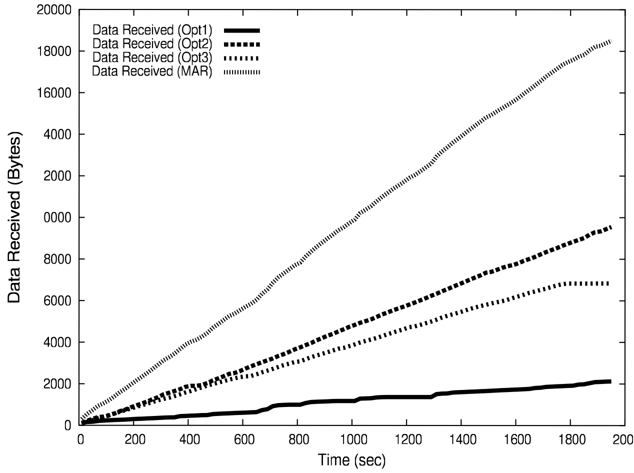

Figure 13: MAR Streaming Performance. Cumulative Data Arrival when using a) the MAR router with three parallel interfaces, and b) each interface stand alone

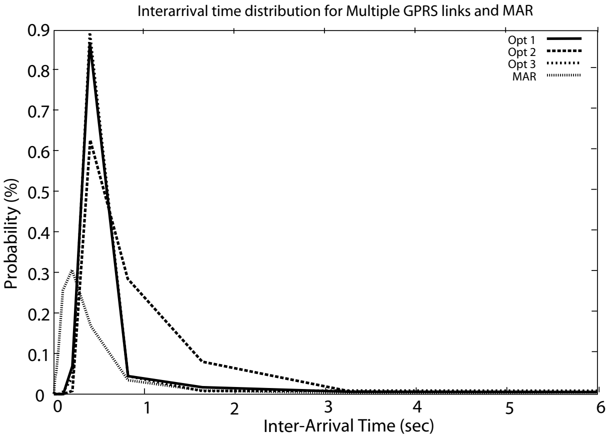

Figure 14: MAR Streaming Performance. Inter-arrival time distribution for a) the MAR router with three parallel interfaces, and b) each interface stand alone

a single interface, end-users suffer frequent buffer starvations, which significantly degrades their streaming experience. However, using the MAR system, even targeting a streaming rate that is four times higher than the one targeted with individual interfaces, the amount of outstanding data is always positive and well over the starvation threshold. Thus, MAR can survive blackout periods and sudden burstiness much better than using any other single interface and therefore provide a much smoother streaming experience.

In the Table 3 we calculate the exact starvation probability for different interfaces and for the MAR system. From this table we can observe that the probability of buffer starvation for a single interface, even when the streaming rate is very low, is quite significant. However, MAR can target streaming rates in the order of 76 kbps with a null probability of suffering an interruption during the 30 minute video clip. Compared to a system that only uses one wireless interface, on average, MAR prevents streaming interruptions 23% of the time while targeting streaming rates that are 4 times higher.

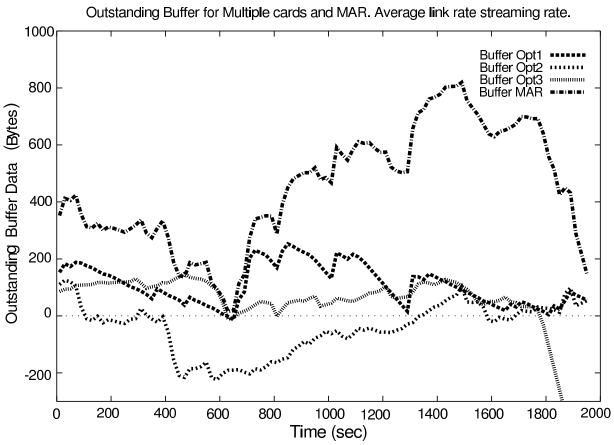

Figure 15: MAR Streaming performance. Outstanding Data buffer for a) the MAR router with three parallel interfaces, and b) each interface stand alone.

6. OPEN ISSUES

As we have discussed, network diversity can be exploited in a wireless overlay and across cellular networks for systems that will use multiple networks simultaneously, such as MAR. Even if we consider that coverage offered by cellular network infrastructures has substantial overlap, MAR's ability to exploit distributed diversity from networks will be for most cases limited to performing good bandwidth aggregation. However, there might also be cases when the MAR router will have to handover sessions across different networks (e.g., GPRS, 3G, and WLANs). This is particularly true for the case when MAR has to exploit network diversity using WLANs. Each network interface in a MAR router will usually be associated with an IP address; however, there will be occasions where network interfaces have less predictable coverage (for example, moving *in-and-out* of 'hotspot' WLANs [28]). For such cases, we need to understand the impact mobility can have on overall (vertical) handover performance. A thorough description of our practical experiences with GPRS–WLAN vertical handovers is available [9].

Attempting to utilise energy efficiently on a MAR router

Table 3: Probability of Buffer Starvation for MAR and each individual interface. 10 sec initial playout buffer. system. Mobile environment. TCP traffic.

Operator	Streaming Rate (kbps)	Starvation Probability
Opt1	8.7	2%
Opt2	39	57%
Opt3	28	12%
MAR	76	0%

is a challenge since the consumption is closely tied to the amount of data sent and the number of active interfaces. Developing an efficient algorithm for managing consumption would be a trade-off in optimising the available link capacity per-user versus minimising the amount of power consumed and ultimately therefore would involve some policy definition. It is envisaged however that energy concerns are not as vital for a router that is transported in a public vehicle such as a bus or train which typically have less constraints on battery power than a small device battery.

MAR can accommodate plug-in schedulers and session-level protocols. Based on the results presented in Section 4.2 we believe that there is room for improving the MAR session-level protocol through the utilisation of techniques such as FEC over large data sets, or through partitioning large objects into smaller blocks. The advantage of such a system would be to help unify the size of all transfer objects such that the effects of link blackouts or reduced throughput are minimised to only small outstanding data transfers. In such circumstances the transfer might even be restarted over a better performing link. Overall, it is anticipated that such an approach should help to create a better load balanced system.

End-to-end IP based security poses some additional challenges to the MAR architecture. Such issues however are very common in networks nowadays which utilise NAT devices to share one or more routable addresses between multiple hosts. The approach typically used to solve the end-to-end IP level communication is dynamic IPSec pass-through which requires the NAT device to additionally keep state of all IPSec requests initiating from within the network and forward the external traffic relating to that request on to the correct host. Higher level security protocols such as Transport Level Security (TLS) and SSL used in secure web and email browsing should be unaffected as long as end-to-end TCP communication can be maintained.

Security at the link level itself for GPRS is fairly robust compared to many technologies due to the Layer 2 SIM-based authentication and, in some cases, further CHAP or PAP based authentication at the PPP level. Once authenticated, the communication over the wireless air interface is encrypted using one of a selection of ciphers which provide adequate protection against eavesdropping or active data attacks.

One aspect of the communication architecture that has not been addressed is the provision of an upstream channel. In the TCP scheduling case, it is clear that the same interface as used for the downlink traffic for a connection should be used for the uplink TCP traffic. This is primarily due to the fact that many cellular providers operate NAT/firewall policies for each connection, and consequently the state must be initiated from the mobile client. In the proxy-server case which utilises a proprietary UDP communication protocol for optimising the link utilisation, the same NAT rules may also apply, however the actual content of the data that is transmitted over each link does not have to correspond directly to the downlink traffic. Consequently a similar upstream scheduling policy could be applied in order to share the load fairly across all interfaces. An additional benefit of this approach is to enable better flow control per interface, and provide faster detection of blackouts.

The economic or pricing model applied by GPRS providers is an interesting benefit to the MAR architecture. Instead of just paying a fixed amount per link as is usual for a data channel in a wired environment, GPRS traffic is typically charged by volume in addition to a standard line rental charge. The benefit therefore to MAR is that the incremental cost of adding more lines, increasing the throughput and consequently the performance of the router but sending the same number of bytes of data across the links is small. This makes the provision of data services to users over a MAR router with multiple links an attractive model since the cost to increase the available throughput is low, whilst users typically expect to pay more for such improvements in service.

The pricing model might further be developed to accommodate policy based routing as determined by a charging metric per byte. In this circumstance, the MAR router would ideally be capable of selecting the cheapest route for low channel usage traffic, and only transmit across the more expensive links when traffic load is high. Accordingly, a dynamic weighted scheduling policy would be required that would bias the assignment of transfers across channels.

Wide-Area Coverage of Operators is potentially a limiting factor of the MAR architecture. In urban, highly populated areas the coverage by operators is sufficiently balanced such that MAR can leverage the benefits of network diversity. In less populated areas, the coverage is more likely to be significantly reduced and to exhibit higher performance correlation between providers. The location and provision of service by providers is generally a trade-off between the cost of maintaining a base station and the anticipated number of users who will benefit from the availability of that station. For wider area coverage over less populated areas, the transmission power can be increased, to reduce the amount of infrastructure required, up to a certain government regulated level, however typically users will experience poorer signal quality and potentially higher contention at individual base stations. The benefits of MAR will therefore be limited in less populated areas. This is an area of further study, and will be addressed in future research.

7. RELATED WORK

Berkeley's BARWAN project made several important observations for wireless overlay networks [2]. The wireless overlay network concept is a way to combine the advantages of wireless coverage while still achieving the best possible bandwidth and latency for mobile devices at any point of time. The objective in MAR is somewhat different, the idea here is to exploit the network diversity not just from wireless overlay networks, but also from the network diversity of the pervasive cellular infrastructure. While the BARWAN project mainly focused on low-latency inter-network handovers between overlays, MAR aims to exploit the network diversity inherent in wireless access to aggregate bandwidth that can be offered as a larger and more stable pipe to the end users. Similarly, the IOTA Project [1] deals with the integration of 802.11 WLANs and 3G Networks. However, their work focuses on how to seamlessly roam across these two networks rather than exploiting diversity across multiple wide-area wireless links.

In some ways, the idea of exploiting network diversity for sustainable data rates in communications channels from wireless overlay and cellular networks has quite similar objectives to that of resilient overlay networks [5]. In RONs, applications use an overlay network to identify good and bad paths and switch from one path to another as necessary, whereas in MAR the client uses the diversity in the wireless access to benefit from many links simultaneously, thereby improving end-to-end reliability and performance. In the same context, [6] proposes to use multiple 802.11 APs to provide low latency video streaming to a single receiver.

MIT's Personal Router (PR) project [4] has a broader objective than MAR. The main idea behind PR is to provide technological infrastructure that supports mobile access to wireless services, along various dimensions such as network support with fast handover, pricing, QoS, network traffic monitoring and user modelling. While the PR Project evaluates many key issues related to wireless access; the main objective in MAR (of exploiting network diversity for reliability and performance) is different from that in PR. However, some innovations in PR might still be applicable for MAR.

Related research projects include the Mobile People Architecture (MPA) [20], the ICEBERG project [19], and the TOPS architecture [21]. All the three projects attempt to provide user level mobility within one or more network types. The MPA uses a person-level router, the Personal Proxy, that tracks a mobile user's location, and accepts communication on the user's behalf, performs any conversions, and then forwards communications to the user. The ICEBERG and TOPS approach depends upon tracking proxy (or tracking router) nodes within the network.

Other research project close to the MAR system is the MOPED project. In [22] MOPED explores efficient inverse multiplexing at the transport level to aggregate multiple wireless channels and differentiate transmission losses from congestion losses in wireless links. Similarly, in [23] MOPED explores how non-conflicting local area wireless technologies can be used to improve throughput and minimize mobility problems. The MAR system builds on top of these concepts and presents extensive results of the performance and diversity of a multi-link bandwidth aggregation system over wide-area wireless links. Wide-area wireless links have a different set of problems compared to local area links (e.g. losses are hidden from the application through link-layer retransmissions, different links may conflict with each other) and, thus, require a careful separate study. To this extent, the GPRSWeb proxy system [11] is a system that efficiently improves end-user performance over wide-area wireless links, however, it does not exploit network diversity by using multiple network interfaces.

A simple approach for a standalone MAR router is to aggregate bandwidth from multiple links is to use link-layer striping techniques [16, 17], where different packets are sent to each interface regardless of their connection id. However, such schemes work quite poorly in wireless links with large performance fluctuations [14]. Other approaches such as PTCP have also been proposed at the transport layer for bandwidth aggregation [14]. However, these schemes rely on the congestion window to be a tight approximation of the available bandwidth-delay product to be able to efficiently stripe different packets in the different interfaces. In many real deployments (e.g. GPRS, CDMA 1xRTT), wireless systems include deep buffers in the Base Station Controllers to mitigate burstiness and therefore artificially inflate the congestion window to larger than the true bandwidth-delay product.

8. CONCLUSIONS

In this paper, we discussed the limitations of the current wireless access systems in the wide-area. To that end, we made a case for exploiting the network diversity in wireless access. We have argued that distributed diversity in wireless access from different wireless overlay and cellular networks can be leveraged to provide a sustainable and reliable wireless communication channel. Based on experiments with production networks, we have shown that there is a substantial overlap in terms of coverage being offered by many of these operators (e.g. Vodafone, Orange etc.) and also across networks (e.g. GPRS, 3G, and/or WLANs).

We introduced MAR, a Mobile Access Router that utilises multiple wireless access links to aggregate bandwidth and provide local users with a smoother, more reliable access network than can typically be provided by a single cellular link. By leveraging the diversity provided through different wireless technologies, different networks provided by individual operators and even diversity through separate channels allocated by the same base station, MAR can provide more stable and sustainable data rates for applications such as web browsing, email and data streaming.

The benefits of employing a simple per-TCP connection scheduler presented, along with analysis showing what additional benefit could be achieved using a more sophisticated striping technique made possible by the server proxy. The possibilities for including other higher-level optimisations in the server proxy were also discussed, such as caching and compression. These benefits can all be achieved without requiring users to perform any software or configuration updates on their mobile devices.

Work on the MAR project is continuing, including a more detailed study of the benefits of utilising different link scheduling algorithms, as well as providing enhanced proxy services through greater collaboration between the MAR server proxy and the MAR router.

9. REFERENCES

[1] M. Buddhikot, G. Chandranmenon, S. J. Han, Y. W. Lee, S. Miller and L. Salgarelli, "Integration of 802.11 and Third Generation Wirelesss Data Networks", *Proceedings of IEEE Infocom 2003*, San Francisco, April 2003.

[2] R. H. Katz and E. C. Brewer, "The Case for Wireless Overlay Networks", *In Proceedings of SPIE Multimedia and Networking Conference (MMNC'96)*, San Jose, CA, Jan 1996.

[3] B. Walke and V. Kumar, "Spectrum issues and new air interfaces", *Computer Communications*, 26(2003), pages 53-65, Elsevier.

[4] D. D. Clark and J. T. Wroclawski, "The Personal Router Whitepaper", MIT Laboratory for Computer Science Technical Report. Version 2.0 – March 2001.

[5] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "The Case for Resilient Overlay Networks", In *Proceedings of HotOS–VIII*, Schloss Elmau, Germany, May 2001.

[6] A. Miu, J. G. Apostolopoulos, W. Tan, and M. Trott, "Low-Latency Wireless Video Over 802.11 Networks Using Path Diversity", In *IEEE International Conference on Multimedia and Expo*, Baltimore, MD, July, 2003.

[7] D. B. Johnson, C. E. Perkins and J. Arkko, “Mobility Support in IPv6” (draft-ieft-mobileip-ipv6-23.txt), May 2003. `http://www.ieft.org`.

[8] N. D. Tripathi, J. H. Reed and H. F. Vanlandingham, “Handover in Cellular Systems”, *IEEE Personal Communications*, December 1998.

[9] R. Chakravorty, P. Vidales, L. Patanapongpibul, K. Subramanian, I. Pratt and J. Crowcroft, “Performance Issues with Vertical Handovers – Experiences from GPRS Cellular and WLAN hot-spots Integration”, In *Proceedings of the IEEE Pervasive Communications and Computing Conference*, March 2004.

[10] R. Chakravorty, S. Katti, I. Pratt and J. Crowcroft, “Flow Aggregation for Enhanced TCP over Wide-Area Wireless”, in *Proceedings of INFOCOM 2003*. `http://www.cl.cam.ac.uk/users/rc277/gprs.html`

[11] R. Chakravorty, A. Clark and I. Pratt, “GPRSWeb: Optimizing the Web for GPRS Links”, In *Proceedings of the ACM/USENIX MobiSys 2003*, May 2003, San Francisco, USA.

[12] P. Rodriguez and V. Fridman, “Performance of PEPs in Cellular Wireless Networks”, In *Proceedings of the 8th Workshop on Web Content Caching and Distribution*, Sep 2003, New York.

[13] P. Rodriguez and E. W. Biersack, “Dynamic Parallel-Access to Replicated Content in the Internet”, In *ACM/IEEE Transactions on Networking*, August 2002.

[14] H.-Y. Hsieh and R. Sivakumar, “A Transport Layer Approach for Achieving Aggregate Bandwidths on Multi-homed Mobile Hosts.” In Proceedings of the ACM MOBICOM, Atlanta, GA USA, September 2002.

[15] M. C. Chan and R. Ramjee, “TCP/IP Performance over 3G Wireless Links with Rate and Delay Variation”, in *Proceedings of ACM Mobicom 2002*, September 2002.

[16] J. Duncanson, “Inverse multiplexing”, *IEEE Communications Magazine*, 32(4):34-41, Apr. 1994.

[17] C. B. Traw and J. Smith, “Striping within the network subsystem”, *IEEE Network Magazine*, 9(4):22-32, July 1995.

[18] J. W. Byers, M. Luby, and M. Mitzenmacher, “Accessing Multiple Mirror Sites in Parallel: Using Tornado Codes to Speed Up Downloads”, in *Proceedings of INFOCOM 1999*, New York, NY, USA.

[19] A. D. Joseph, B. R. Badrinath and R. Katz, “The Case for Services over Cascaded Networks”, In *Proceedings of WOMMOM’98*, 1998.

[20] M. Roussopoulos, P. Maniatis, E. Swierk, K. Lai, G. Appenzeller, M. Baker, “Person-Level Routing in the Mobile People Architecture”, In *Proceedings of the USENIX Symposium on Internet Technologies and Systems*, October 1999.

[21] N. Anerousis, et. al., “TOPS: An Architecture for Telephony Over Packet Networks”, In *IEEE Jounal of Selected Areas of Communications*, 17(1), January 1999.

[22] L. Magalhaes and R. Kravets, “End-to-End Inverse Multiplexing for Mobile Hosts” In *the Journal of the Brazilian Computer Society.*

[23] C. Carter and R. Kravets “User Devices Cooperating to Support Resource Aggregation” In *The 4th IEEE Workshop on Mobile Computing Systems and Applications*, 2002.

[24] T. Nguyen and A. Zakhor “Distributed Video Streaming over the Internet” In *Multimedia Computing and Networking (MMCN)*, 2002.

[25] T. Nguyen and A. Zakhor “Distributed Video Streaming with Forward Error Correction” In *Packet Video*, 2002.

[26] P. Rodriguez, R. Chakravorty, Julian Chesterfield, Ian Pratt, Suman Banerjee “Diversity in the MAR system”, Microsoft Research– Technical Report, Sep 2003. `http://www.research.microsoft.com/~pablo/papers/all_fig.ps`

[27] “Squid Proxy Cache” `http://www.squid-cache.org`.

[28] J. Ott and D. Kutscher “Drive-thru Internet: IEEE 802.11b for Automobile Users” In *Proceedings of Infocom*, 2004.