Experiences with Heterogeneous Wireless Networks, Unveiling the Challenges

Pablo Vidales, Leo Patanapongpibul, Glenford Mapp, Andy Hopper

Laboratory for Communication Engineering, University of Cambridge,
JJ Thompson Avenue, Cambridge CB3 0FD, U.K.
Emails: {pav25, lbp22, gem11, hopper}@eng.cam.ac.uk
http://www-lce.eng.cam.ac.uk/

Abstract

The integration of heterogeneous wireless networks is one of the most important features that are needed for the future deployment of new wireless technologies such as 4G mobile systems. However, the designers of these systems are also committed to using Internet Protocol (IP) as their core network infrastructure. Thus, support for vertical handover, location, and mobility management need to be provided within the IP framework. This paper looks at the development of a wireless testbed to study such issues.

1. Introduction

We are witnessing the development and deployment of a large number of wireless networking technologies including 3G, WLANs, Bluetooth, and Ultrawideband. At the same time we are seeing a convergence of core networking infrastructure on the Internet Protocol Suite (IP) [5]. IPv4 is widely deployed throughout the Internet and there is now a serious attempt to deploy IPv6, which offers a much large IP address space (128 bits) as well as easier protocol processing and mobility enhancements.

Looking at this from an OSI perspective, Layers 1 and 2 are being hugely expanded in the wireless domain with Layer 3, the network layer, becoming fixed on IP. The result is that IP must support key features associated with wireless networking including handover, location, and mobility management.

Moreover, due to the multiplicity of choices available from many cellular/wireless network providers, different access technologies, and disparate application's QoS requirements, there is a significant need to come up with a single unified approach. The 4G architecture envisions highly flexible and adaptive integration of diverse mobile client systems and network technologies to support built-in capability for seamless wireless access.

Implicitly, this also means that there will be a need for mobile devices that can cope with the complexity and dynamics of next generation (4G) wireless access environments. With more technologies, services, and devices joining the fray, we can expect that the gap between the service levels offered by new access networks will close, adding more complexity to the handover process (see table 1).

IP does not support an official handover mechanism and present commercial handover policies are inadequate in dealing with vertical handovers. In addition, there needs to be support for integrated location management for seamless roaming, as well as for micro- and macro-mobility management. Mobile IP has been developed to support mobility. Mobile IPv4 [10] has been deployed for sometime while MobileIPv6

Network	Coverage	Data Rates	Mobility	Cost
Satellite (B-GAN)	World	Max. 144 kb/s	High	High
GSM/GPRS	Aprox. 35 Km	9.6 kb/s up to 144 kb/s	High	High
IEEE 802.16a	Aprox. 30 Km	Max. 70 Mb/s	Low/Medium	Medium
IEEE 802.20	Aprox. 20 Km	1-9 Mb/s	Very high	High
UMTS	20 Km	up to 2 Mb/s	High	High
HIPERLAN 2	70 up to 300 m	25 Mb/s	Medium/high	Low
IEEE 802.11a	50 up to 300 m	54 Mb/s	Medium/high	Low
IEEE 802.11b	50 up to 300 m	11 Mb/s	Medium/high	Low
Bluetooth	10 m	Max. 700 kb/s	Very low	Low

Table 1. Diversity in existing and emerging wireless technologies demand flexible and adaptive roaming devices.

has just recently been made an RFC [6]. However,we believe that better mobility management schemes are needed to support this new environment.

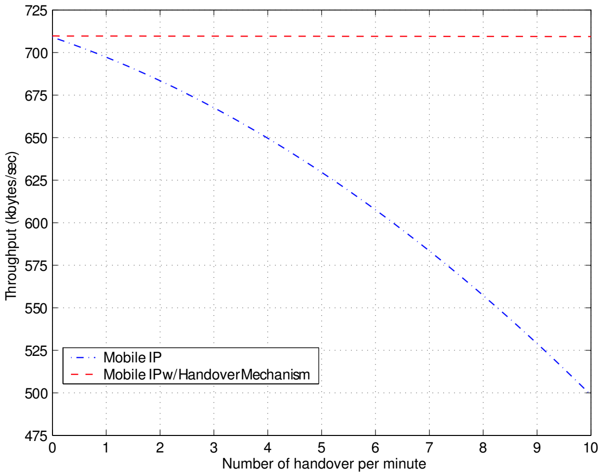
In 2003, the departments of the Laboratory for Communication Engineering (LCE) and the Computer Laboratory (CL) at the University of Cambridge came together to develop a testbed to study these issues (Figure 3). This paper presents the design, integration and development of the testbed. Recent results are presented and past, current and future work are discussed. The goal of these efforts is to produce a platform that fully integrates heterogeneous wireless technologies because in the near future mobile devices will have several wireless interfaces and users will expect connections to be seamlessly managed. Future 4G networks will have such features so the testbed can be regarded as an early attempt to build such a system.

The paper is therefore structured as follows: Section 2 looks at handover mechanisms and introduces the concept of client-based handover. Section 3 looks at vertical handovers and introduces PROTON, a policy mechanisms to decide how, when, and where vertical handovers occurs. Section 4 introduces the Cambridge Wireless Testbed and recent latency results for vertical handovers are presented. We discuss future work on providing better mobility support within the testbed in Section 5. Related work is included in Section 6. To conclude in Section 7, we summarise the work done.

2. Handover mechanisms

Handover introduces packet loss and latency which can severely damage data communication. TCP, part of the IP Suite and used throughout the Internet, reacts extremely adversely to packet loss. This is because TCP assumes all packet loss is due to congestion and limits its window-size as a result, leading to reduced performance. Multimedia applications which normally have strict temporal constraints can be also affected by increased latency.

Handover mechanisms must therefore be geared to minimising these effects and maintaining good transport performance. On the control plane, handover mechanisms can be divided into two types. Network-controlled handover happens when the network is responsible for the handover. Here, the mobile node reports the received signal strengths from various base stations to the network which decides when to switch the node to another attachment point. Thus, to minimise the effects of handover various buffering schemes have been proposed, which attempt to buffer packets during handover. However, the approach is unable to determine the best time to handoff as it does not have enough information about what is happening in the mobile node context. In addition, network-controlled handover is unsuitable for vertical handovers, since each network will have to be aware of the characteristics of all the other wireless networks to be able to take the decision.



MIPv6 Enhancement	Handover Latency (seconds)				Average Ratio
	Min	Average	Max	Stdev	
None	2.277	4.209	4.759	1.041	0.165
RA Cache	0.097	0.397	1.042	0.430	0.019
Scenario 1	0.112	0.633	0.895	0.241	0.029
Scenario 2	0.191	0.238	0.258	0.024	0.011

Right: Throughput of a UDP data stream, from correspondent node (i.e. http server) to the mobile node, versus horizontal handover frequency.

Left: The handover latency during a 10Mb file download from a HTTP server performed over 10 runs. In *Scenario 1* the handover occurs between two non-overlapping base stations. On the contrary, in *Scenario 2* handover was forced between two overlapping base stations.

Figure 1. Experiments on the effectiveness of the client-based handover mechanism

2.1 Client-based handover mechanism

It was therefore decided to take an end-system approach to solving handover issues, in which the client –not the network– controls the handover. This means that the mobile node can take into account various factors such as the state of transport connections, the applications running on the mobile, the physical context, and the traffic conditions in all available overlays, so that handover is performed at the optimum time thus maintaining the best possible level of service. The client-based handover mechanism provides the mobile node with a method to tackle the following issues: *control and force handovers*, *determine the best link*, *handover at the appropriate time*, and *resume active TCP connections*.

The combination of these methods is a novel way to assist Mobile IP's movement detection. As a result of the above issues, the client-based handover mechanism incorporates a cache to store IPv6 Router Advertisements from nearby network access points and Home Agents. RAs are stored in a cache, called the *RA Cache*, along with parameters such as signal strength, link-layer metrics, and the time at which the RA was captured. When the signal from the access point to which the mobile node is currently attached falls below a given threshold, the RA Cache is examined to determine the next access point to which the mobile node should be attached. Handover needs to be performed at the appropriate time, thus any active TCP connections also need to be considered. Handover is then forced to the next point of attachment while any active TCP connections are forced to resume immediately. This technique to swiftly resume active TCP connections avoids TCP slow start and packet loss. Details of the technique are discussed in the next section.

Experiments on the effectiveness of the client-based handover mechanism were carried out on the testbed for cross domain horizontal handovers between IEEE 802.11b wireless LAN base stations. The performance of the client-based handover mechanism is shown in Figure 1 (left side) where the throughput of a UDP data stream from a correspondent node to a mobile node is plotted versus the handover frequency of the mobile node. These results are obtained from a scenario where the wireless coverage area of two base stations do not overlap. This scenario is called a *non-continuous handover*.

In these experiments, the boundaries of the network coverage areas are immediately adjacent to each other with the smallest possible gaps. This is necessary to obtain reliable results from the testbed for experiments without the client-based handover mechanism since Mobile IP movement detection is triggered every time a RA is received by the mobile node. Note that the signal strength at the network coverage area boundary is maintained at a high level to prevent a drop in throughput resulting from a weak link connectivity or an increment in bit error rate.

One of the lines in Figure 1, where the client-based handover mechanism is built into the mobile node, shows the throughput to be virtually constant despite the increase in the handoff rate. In the second line where client-based handover mechanism is not implemented, the throughput decreases with higher handoff frequencies. This is largely due to the delay in discovering the link change. The client-based handover mechanism avoids this latency component by its ability to detect the new link immediately and, as a result, to spontaneously request for a RA by sending a Router Solicitation to the access router on the new link. If a RA of the new link happens to be cached in the node, then the mechanism will process it instead.

Handover experiments on a second scenario where the network coverage area of two or more base stations overlap (called a *continuous handover*) shows the throughput to be similar for cases with and without the client-based handover mechanism.

2.2 Resuming active TCP connections

In order to maintain good transport performance, the mobile node keeps a copy of the last outgoing or last incoming TCP ACK packet. At least three copies of these ACKs are retransmitted to the sender or inserted to the mobile node's incoming TCP buffer to trigger the TCP fast retransmit algorithm and to immediately resume any active TCP connections after a continuous handover. This method works well when the node has a clear set of available nodes to which to handoff.

However, in the non-continuous handover case, it is better to stop the sender from transmitting until another point of attachment is found. This is done by setting a TCP receive window size of zero in the last outgoing or last incoming TCP ACK packet before the mobile node performs a handover. This halts the transmission of any more TCP data. When a new point of attachment is found the mobile node then advertises a non-zero TCP receive window size causing its TCP connections to be resumed.

The table in Figure 1 shows a summary of the horizontal handover latencies from experiments carried out on the testbed. The handover is performed between IEEE 802.11b wireless LAN base stations. From the table (right side), *Scenario 1* is the non-continuous handover case and *Scenario 2* is the continuous handover case. The handover latency results for *None* and *RA Cache* are obtained from both Scenarios 1 and 2. We can see from the average handover latency and download time ratio in the table that there is a significant reduction in the handover latency when the client-based handover mechanism without the TCP enhancements (*RA Cache*) is enabled as compared to base Mobile IP (*None*). This result is not surprising knowing that the Mobile IP movement detection implementation used in the testbed is based on network-layer triggers, i.e. RAs. However, when we compare the handover latency of the *RA Cache* to the *Mechanism* under Scenario 2, there is a 41% reduction in the handover latency.

The handover latency of the *Mechanism* under Scenario 1 in the table does not include a subnetwork outage period. The values obtained are the effective handover latency times. The results show a significantly lower latency compared to the base Mobile IP. This suggests that this handover method where the sender is forced to stop transmitting any data by the mobile node during a handover could also be used in Scenario 2. Nevertheless, the client-based handover mechanism completely avoids TCP slow start in both Scenarios.

3. Vertical Handover Mechanisms

The mechanisms presented above have been expanded to support vertical handover. Several techniques were evaluated to minimise the handover latency between different networks. These mechanisms are detailed below:

Fast Router Advertisement: This technique can improve handover latency, by reducing detection time. However, there is a trade off between the handover latency and bandwidth: *router advertisement frequency should be reduced as we move up in the overlay model.* Decreasing router advertisement period to very low values (40ms-70ms) as that typically specified in the latest IETF Mobile IPv6 RFC [6] can incur substantial overhead –up to 45% of the available bandwidth in a GPRS network– in the top overlays (e.g., satellite, GPRS, and UMTS).

Router Advertisement caching: This method also avoids detection time. It is an adaptation of the solution presented in Section 2, now for vertical handovers. In this scheme, RAs are cached *a priori* by the mobile

Upward handover	Without BU bi-casting	With BU bi-casting	Reduction
LAN to WLAN	7.5ms	1.9ms	75 %
WLAN to 3G	750ms	156ms	79.2 %
3G to GPRS	2500ms	1000ms	60 %
WLAN to GPRS	2500ms	506ms	79.76 %

*R_t is calculated pondering the following RTT values: *LAN=0.2ms*, *WLAN=3 ms*, *3G=300ms* (expected value), and *GSM/GPRS=1000ms*.

The table shows the reduction in Registration Time (R_t) for upward vertical handovers (lower values), using BU bi-casting. The Registration process time period (for Mobile IP) is given by:

$$R_t >= RTT_{SmallerLatency} + (N * \frac{RTT_{SmallerLatency}}{2})$$

where N is the number of correspondent nodes

Figure 2. Mobile IP specifies to use the new attachment point during the registration process. BU bi-casting mechanism reduces R_t and minimises overall vertical handover latency with minimum overhead.

node, however, it only works for *upward vertical handovers* due to the fact that only the RAs from the above overlays are available. Thus, when the handover decision occurs, the detection time for RA lookup during handover execution is eliminated, improving overall performance.

Binding Update bi-casting: The minimum limit for vertical handover latency is given by the latency in the corresponding channel –which can be observed using the Round Trip Time (RTT) of the network. Mobile IPv6 specifies the use of the *new attachment point* to send the signalling for the registration process. We propose to bi-cast Binding Updates, thus the registration time is limited by the smallest RTT and not by the latency of the new network, which can be higher than the current network (see Figure 2).

3.1 Policy-based handover solution

For homogeneous wireless networks, handover is a well-understood technology. For heterogeneous wireless networks, published vertical handover research has focused almost exclusively on simple mechanics of inter-working between different signalling protocols. There has been a dearth of attention given to the algorithms and policies that control the access and handover decisions, and the related performance and infrastructure implications.

Policies are rules that govern the choices and behaviour of a system. A considerable amount of work has been done to use policies as a tool to create flexible and highly adaptive end systems [11]. With PROTON, we aim to provide complete mobility support based on context knowledge represented as policies that optimise the use of bandwidth and maximise user experience in 4G wireless networks.

We consider that a policy-based solution is in concordance with our aims which are to provide a set of dynamically changing policies to help 4G mobile devices to adapt without incurring huge delays. PROTON will provide flexibility and adaptiveness to 4G users while coping with dynamics and heterogeneity in the environment.

The main idea is to work toward a policy-based support system in which context fragments are collected and expressed. With this solution, multi-mode mobile devices receive assistance during vertical handovers to seamlessly perform informed and accurate decisions based on context knowledge.

Current handover algorithms rely on simple decisions based on signal strength, and the most sophisticated perform handovers considering only connection state and traffic type. Although this might be appropriate for homogeneous environments, we consider that mobile users roaming between various technologies need a more complete solution to decide where, when, and how to perform the handover. PROTON is a three-layered middleware: *Context Management layer*, *Policy Management layer*, and *Enforcement layer*, and it sits on top of the network layer (the detailed solution is available in [12]).

4. Testbed

To closely emulate the next generation (4G) integrated networking environment, our experimental testbed setup consists of a tightly-integrated, Mobile IPv6-based GPRS-WLAN-LAN testbed as shown in Figure 3.

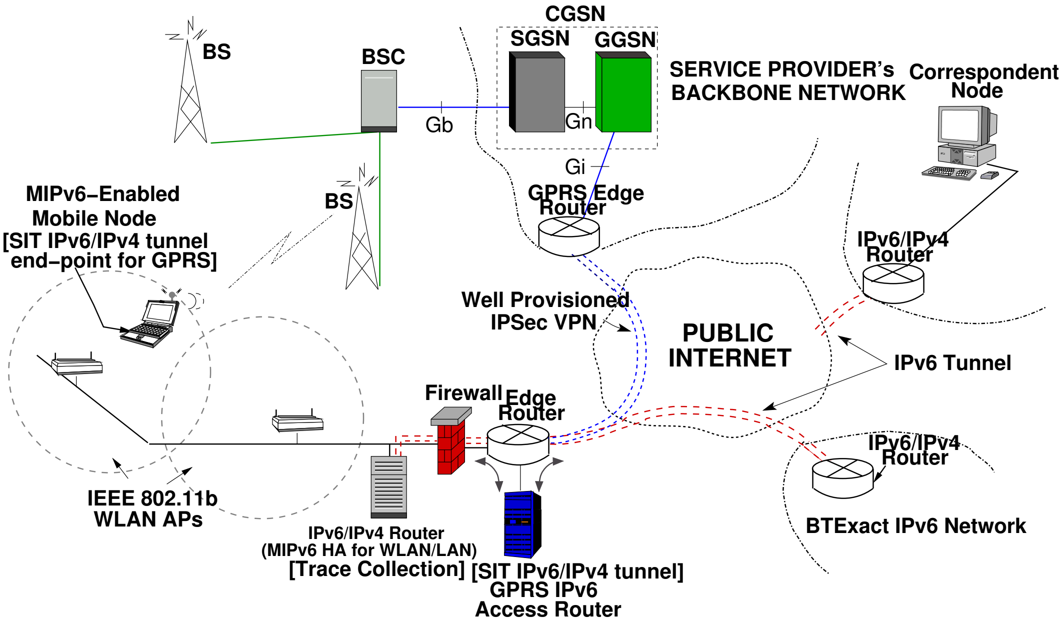


Figure 3. A tightly-integrated architecture includes the possibility of seamless handovers between different access networks, and foresees the need include mobility support features in the Internet Protocol Suite

The cellular GPRS network infrastructure currently in use is the Vodafone UK's production GPRS network. The WLAN access points (APs) are IEEE 802.11b APs. Our testbed has been operational since March 2003, and results showing how we optimise vertical handovers are detailed in [3].

In the testbed, the GPRS infrastructure comprises base stations (BSs) that are linked to the SGSN (Serving GPRS Support Node) which is then connected to a GGSN (Gateway GPRS Support node). In the current Vodafone configuration, both SGSN and GGSN node are co-located in a single CGSN (Combined GPRS Support Node). A well provisioned virtual private network (VPN) connects the Lab network to that of the Vodafone's backbone via an IPSec tunnel over the public Internet. A separate "operator-type" RADIUS server is provisioned to authenticate GPRS mobile users/terminals and also assign IP addresses.

For access to the 4G integrated network, mobile nodes (e.g., laptops) connect to the local WLAN network and also simultaneously to GPRS via a Phone/PCcard modem. The mobile node's MIPv6 implementation is based on that developed by the MediaPoli project [8], chosen for its completeness and open source nature. We brokered a semi-permanent IPv6 subnet from BTExact's IPv6 Network, which connects us to the 6BONE. Using the address space, we are able to allocate static IPv6 addresses to all our IPv6 enabled mobile nodes. A router in the lab acts an IPv6/IPv4 tunnel end-point to the BTExact's IPv6 network. This router is also There is an IPv6 access router (Home Agent) for the lab's fixed-internal IPv6-enabled network and also for internal WLANs (shown in Figure 3). Routing in the Lab has been configured such that all GPRS/WLAN user traffic going to and from mobile nodes are allowed to pass through the internal router, enabling us to perform traffic monitoring.

Since the GPRS cellular network currently operates only on IPv4, We use a SIT (Simple Internet Translation) to tunnel all IPv6 packets as IPv4 packets between the mobile node and a machine providing IPv6-enabled access router functionality on behalf of the GPRS network. Ideally, the GGSN in the GPRS network would provide this functionality directly, but using the tunnel incurs only minor overhead.

Using the testbed, we have evaluated the impact that vertical handovers can have on the Internet Protocol Suite (detailed study on performance is presented in [3]). From tables in Figure 4, we can understand the need to modify current protocols to support mobility management, location, and handovers in future environments without incurring in huge delays and overheads. Some latency values, using the current Mobile IP specifica-

WLAN ⇒ GPRS	Mean	Std. Dev.	Min.	Max.
Detection time (t_d)	808	320	200	1148
Configuration time (t_c)	1	0	1	1
Registration time (t_r)	2997	416	2339	3649
Total handover latency (t_h)	3806	327	3323	4438

GPRS ⇒ WLAN	Mean	Std. Dev.	Min.	Max.
Detection time (t_d)	2241	968	739	3803
Configuration time (t_c)	1	0	0	1
Registration time (t_r)	4654	1698	2585	7639
Total handover latency (t_h)	6897	1178	5322	8833

LAN ⇒ GPRS	Mean	Std. Dev.	Min.	Max.
Detection time (t_d)	1168	460	347	2070
Configuration time (t_c)	1	0	1	1
Registration time (t_r)	3307	585	2299	4759
Total handover latency (t_h)	4476	520	2806	5107

GPRS ⇒ LAN	Mean	Std. Dev.	Min.	Max.
Detection time (t_d)	2058	1030	1	3257
Configuration time (t_c)	1	0	1	1
Registration time (t_r)	4466	1449	2357	7183
Total handover latency (t_h)	6525	1229	4011	8197

Figure 4. Latency partition for vertical handovers during a TCP transfer: *upper left* **WLAN → GPRS,** *upper right* **GPRS → WLAN,** *lower left* **LAN → GPRS,** *lower right* **GPRS → LAN.**

tion, are unacceptable to offer satisfactory QoS to mobile users (e.g., upward handovers *3.8s* and downward handovers *6.8s* during a TCP transfer between the mobile node and the correspondent node (i.e. http server).

5. Mobility Management for Heterogeneous Wireless Networks

Mobile IP is used for mobility management in IP. However, we believe that it needs to be improved to manage this new environment where a mobile may be simultaneously attached to several wireless interfaces. Work in this area will be concentrated on three fronts.

Naming: Communication over Internet has evolved to mean communication between network interfaces. So each network interface is given an IP address. This has led to the multi-homed problem for devices with several interfaces. We propose to introduce a unique identifier called the Node ID that represents the object during its lifetime and is independent of its network interfaces.

Addressing: Here we will look at the IP address format. The key weakness of current IP addressing is that an IP address identifies the object (more specifically the network interface) being contacted and is also used by the network to route data to the object. This means that when the object changes location a completely new address must be issued but the machine must also be associated with a home address to allow communication with other devices. Thus, as used in Mobile IP, two IPv6 addresses are necessary to support mobile nodes in a wireless network. This solution becomes impractical when devices have several interfaces, each requiring a different home address. We are therefore proposing to split the IP address into a Node ID and a Location ID. The Node ID stays the same but the Location ID changes as the object moves around. Hence an object can have several locations IDs as it may be simultaneously connected to different wireless networks.

Support for Mobility and Location: Since objects may be continuously on the move, there should be more assistance to indicate how mobile an object currently is on a given network and therefore the validity of the Location ID field used to contact the object. We propose to introduce a mobility indicator/hint into the IP format. The value of this indicator changes as the mobile device moves around a given network.So when deciding to sent data to a mobile, the sender will take into account the mobility of the mobile on different networks and may choose the network in which the mobile node appears to be least mobile to avoid the possibility of packet-loss or delay due to handover in networks where the mobile is highly mobile.

6. Related Work

IP technology growth explosion –mainly due to the popularity of Internet services– brings to the fore *network convergence* as an immediate challenge. Thus, considering an IP core network as the next generation architecture, Mobile IP [6] represents a *de facto* solution for macro-mobility, and Cellular IP [2] could be considered a sensible option for micro-mobility.

A number of strategies to perform effective handovers in heterogeneous systems have been explored since 1998, when the concept of *Overlay Networks* first appeared [4]. As part of Daedalus project [4], Helen J. Wang [13] employs schemes based on policies to evaluate the performance offered by each network, and select the most appropriate according to user's criteria. This is probably the first attempt to define a complete mobility solution for inter-system handovers.

This achievement was followed by other policy-based approaches to tackle different handover related problems such as data-flow based selection of the most appropriate access technology [7] and handover initiation [1].

Although many solutions have been proposed to solve inter-system handover challenges, it was only lately that complete mobility support solutions were envisaged. Whereas some groups believe in network-assisted solutions [9], our approach foresees a mobile-based middleware to support users in handover initiation, network selection, handover execution, security, and data adaptation –along the complete handover process.

7. Summary and Conclusions

This paper has described the design and deployment of a wireless testbed to examine issues in the integration of heterogeneous wireless networks using the IP networking framework. We have shown a client-based mechanism based on using IPv6 features which have been extended to handle vertical handovers. A policy-based mechanism called PROTON has also been discussed and work is beginning to look at improving mobility management mechanisms for heterogeneous wireless networks.

Acknowledgements

The authors would like to thank Dr. Frank Stajano, Calicrates Policroniades, and Rajiv Chakravorty for their helpful comments. Pablo Vidales has a scholarship from the Mexican Government through the National Council of Science and Technology (CONACyT). Glenford Mapp is a Visiting Fellow in the LCE and is also a Principal Lecturer in Computer Networks at Middlesex University in North London.

References

[1] S. Aust, D. Proetel, N. A. Fikouras, C. Pampu, and C. Gorg. Policy based Mobile IP handoff decision (POLIMAND) using generic link layer information. In *Proceedings of 5th IEEE International Conference on Mobile and Wireless Communication Networks (MWCN 2003)*, October 2003.

[2] A. Campbell, J. Gomez, C.-Y. Wan, S. Kim, Z. Turanyi, and A. Valko. Cellular ip. Internet Draft (draft-ietf-mobileip-cellularip-00.txt), , Work in Progress, January 2000.

[3] R. Chackravorty, P. Vidales, I. Pratt, and J. Crowcroft. On TCP Performance during Vertical Handovers: Experiences from GPRS-WLAN Integration. In *Proceedings of The Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, March 2004.

[4] Daedalus wireless research group. `http://daedalu.cs.berkeley.edu/`.

[5] ISI-USC. Transmission Control Protocol and Internet Protocol specifications, IETF *RFC 791* and IETF *RFC 793*, 1981.

[6] D. B. Johnson, C. E. Perkins, and J. Arkko. Mobility Support in IPv6 (RFC 3775) `http://www.ieft.org`, 2004.

[7] K. Lai, M. Roussopoulos, D. Tang, X. Zhao, and M. Baker. Experiences with a mobile testbed. In *Proceedings of The Second International Conference on Worldwide Computing and its Applications (WWCA'98)*, March 1998.

[8] Mobile IP for Linux (MIPL) Implementation by HUT Telecommunications and Multimedia Lab, `http://www.mipl.mediapoli.com`.

[9] K. Murray, R. Mathur, and D. Pesch. Intelligent access and mobility management in heterogeneous wireless networks using policy. In *ACM 1st International Workshop on Information and Communication technologies*, pages 181–186, 2003.

[10] C. E. Perkins. IP Mobility Support for IPv4 (RFC 3344) `http://www.ieft.org`, 2002.

[11] M. Sloman and E. Lupu. Security and management policy specification. In *IEEE Network*, volume 16, issue 2, pages 10–19, March/April 2002.

[12] P. Vidales, R. Chakravorty, and C. Policroniades. PROTON: A Policy-based Solution for Future 4G devices. In *Proceedings of 5th International Workshop on Policies for Distributed Systems and Networks (IEEE POLICY 2004)*, New York, United States, June 2004.

[13] H. J. Wang. Policy-enabled handoffs across heterogeneous wireless networks. Technical Report CSD-98-1027, 23, 1998.