HANDOVERS IN WIRELESS OVERLAY NETWORKS

by

Meltem Yıldırım

B.S. in Computer Engineering, Istanbul Technical University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2006

HANDOVERS IN WIRELESS OVERLAY NETWORKS

APPROVED BY:

Assoc. Prof. Fatih Alagöz (Thesis Supervisor)	.....
Prof. Emin Anarım	.....
Assist. Prof. Tuna Tuğcu	.....

DATE OF APPROVAL: 24.05.2006

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Assoc. Prof. Fatih Alagöz for his guidance, support and encouragement during my study. He is my inspiration and role model in achieving so much in such a short lifetime.

I also thank my colleagues in Naryaz Software for their tolerance, cooperation and joy. This teamwork, this synergy propels us in every project.

I especially thank my friends, Ayşegül Tüysüz and Reha Kamil Gerçeker for their nourishing ideas and support. Without their assistance, this thesis would be incomplete.

Finally, I thank my family for their unconditional love, trust, appreciation and help throughout my life.

This thesis is partially supported by the State Planning Organization of Turkey. Project: Next-generation Satellite Networks and Applications, DPT-03K120250.

ABSTRACT

HANDOVERS IN WIRELESS OVERLAY NETWORKS

Wireless Overlay Networks (WONs) are multi-layer networks in which mobile nodes (MNs) move freely within cells of the same overlay or between cells of different overlays. Such a structure provides global access to MNs regardless of time and location. In spite of their advantages, WONs pose many challenges due to the simultaneous existence of *horizontal* handovers between cells of the same wireless network and *vertical* handovers between different overlay networks. Minimizing power drain, handover latency, overhead due to extra messaging during handovers, eliminating the number of unnecessary handovers and maximizing throughput are some of these challenges. While dealing with all this, end users should be isolated from making additional configurations and handovers should be performed seamlessly. Thus, an efficient mobility management system achieving all these goals needs to be built for the WON architecture.

Currently, there are many efforts in dealing with the problems mentioned above. Related work in this area focuses on the development of new mobility management architectures as well as improving the handover decision phase to quickly and accurately trigger handovers without triggering any unnecessary ones. The thesis includes a survey of present handover schemes proposed for WONs and points out their drawbacks. Then, by taking into consideration the drawbacks and incompleteness of the current schemes in literature, this thesis proposes a new mobility management scheme for WONs. The proposed scheme divides into two main parts.

The first part covers the problem of horizontal handovers and presents a solution which reduces latency as much as possible. The proposed model also deals with

oscillations as MNs moves back and forth between two *Access Routers (ARs)*. The simulation results show that the proposed horizontal handover model reduces latency significantly especially in highly crowded subnets at the expense of slightly increased overhead.

The second part covers the problem of vertical handovers. This scheme integrates a *Handover Decision Manager (HDM)* to detect network condition changes in a timely and accurate manner, and a *Subnet Agent (SA)* to maintain the connection and manage handovers. The proposed HDM algorithm considers many factors such as the candidate networks' cost, coverage, data rate and traffic load. The algorithm gives accurate and timely decisions and considerably decreases MNs' power consumption at the expense of slightly increased complexity. The simulations, which compare the proposed HDM scheme with two-level thresholding and utility-based schemes, show that the number of unnecessary handovers and MNs' power consumption are decreased considerably while MNs' throughput is not significantly affected by the proposed algorithm. Meanwhile, the novel SA-based architecture brings a new signaling flow for downward and upward vertical handovers. The performance of the architecture is evaluated by comparing its signaling flow with that of the two latest architectures in literature. The analytical work done for this evaluation shows that the proposed SA-based architecture has a significantly lower overhead and an acceptable latency.

ÖZET

KABLOSUZ ÜSTÜNE BİNDİRMELİ AĞLARDA ELDEĞİŞTİRMELER

Kablosuz Üstüne Bindirmeli Ağlar (Wireless Overlay Network - WON), mobil düğümlerin içinde yatay veya dikey eldeğiştirmeler yaparak serbestçe dolaşabildiği çok katmanlı ağlardır. Bu sayede, mobil kullanıcının konumuna ve sahip olduğu yerel kablosuz bağlantıya göre mümkün olan en iyi bağlanabilirlik sağlanmış olur. Sağladığı yararların yanısıra, bir WON birçok problemi de beraberinde getirir. Bu problemler hem aynı kablosuz ağın hücreleri arasında gerçekleşen *yatay* eldeğiştirmeler, hem de farklı katmanlardaki ağlar arasında gerçekleşen *dikey* eldeğiştirmeler yüzünden kaynaklanmaktadır. Enerji tüketimini, eldeğiştirme gecikmesini, eldeğiştirme sırasında yapılan mesajlaşmayı en aza indirmek, gereksiz eldeğiştirmeleri engellemek ve ağda dolaşan veri hacmini mümkün olduğu kadar arttırmak bu problemlerden bazılarıdır. Ayrıca, tüm bu problemlerle başa çıkarken, son kullanıcıların mobil cihazlarında ekstra ayarlar yapması gerekmemeli ve eldeğiştirmeler son kullanıcının kesintisiz iletişimini etkilememelidir. Kısacası, bir WON için tüm bu problemlerin üstesinden gelebilecek etkili bir *Hareket Yönetim Sistemi* gerekmektedir.

Şu anki durumda, yukarıdaki problemlerle başa çıkabilmek için çeşitli çabalar mevcuttur. Bu konuda yapılan çalışmalar, yeni hareket yönetim sistemlerinin tasarlanmasına ve eldeğiştirme sürecindeki karar verme aşamasının eldeğiştirmeleri hızlı ve doğru bir şekilde tetikleyecek yönde iyileştirilmesine odaklanmıştır. Bu tez çalışması, şu ana kadar WON için önerilen yatay ve dikey eldeğiştirme sistemlerini incelemekte ve bu sistemlerin eksiklerini ortaya çıkarmaktadır. Daha sonra, incelenen sistemlerin eksiklik ve sakıncaları göz önüne alınarak WON yapısı için yeni bir Hareket Yönetim

Sistemi önerilmektedir. Mobil kullanıcıların kablosuz üstüne bindirmeli ağlar arasında serbestçe dolaşabilmesini sağlayan bu yeni sistem iki ana bölümden oluşmaktadır.

Birinci bölüm yatay eldeğiştirmeleri kapsamakta ve eldeğiştirme gecikmesini azaltacak bir çözüm sunmaktadır. Ayrıca, önerilen bu model mobil kullanıcıların iki yönlendirici (Access Router - AR) arasında ileri geri hareket etmesiyle oluşabilecek sinyal seviyesindeki salınımları gözardı ederek gereksiz eldeğiştirmeleri de engellemektedir. Simülasyon sonuçları önerilen yatay eldeğiştirme modelinin özellikle çok kalabalık ağlarda eldeğiştirme gecikmesini önemli ölçüde düşürdüğünü göstermektedir.

İkinci bölüm ise dikey eldeğiştirmeleri kapsamaktadır. Bu model, eldeğiştirme tetiklenmelerine yerinde ve zamanında karar verecek şekilde tasarlanmış *Eldeğiştirme Karar Yöneticisi (Handover Decision Manager - HDM)* ile eldeğiştirme sırasında kullanıcının iletişiminin kesintiye uğramasına engel olan *Alt Ağ Temsilcisi (Subnet Agent - SA)* bileşenlerinden oluşmaktadır. Ayrıca, bu çalışmada, her mobil iletişim cihazına yerleştirilmesi önerilen HDM bileşeni için yeni bir algoritma da sunulmaktadır. HDM algoritması aday ağların maliyeti, kapsama alanı, veri hızı, yükü gibi birçok etmeni gözönünde bulundurmaktadır. Algoritma doğru ve zamanında eldeğiştirme kararları vererek mobil cihazın enerji tüketimini büyük ölçüde azaltmaktadır. Önerilen algoritmayı İki Seviyeli Eşik Sistemi (two-level thresholding) ve Fayda-Odaklı Sistem (utility-based) ile karşılaştıran simülasyonlar sonucunda, gereksiz eldeğiştirmelerin ve mobil cihazdaki enerji tüketiminin azaldığı, buna karşın veri hacminin önemli ölçüde etkilenmediği görülmektedir. Son olarak, yeni SA tabanlı mimari yukarı ve aşağı yöndeki eldeğiştirmeler için farklı mesajlaşmalar önermektedir. Önerilen mesaj akışları, son yıllarda önerilen iki yeni mimarinin mesaj akışlarıyla karşılaştırıldığında SA tabanlı mimarinin kabul edilebilir bir eldeğiştirme gecikmesine sahip olduğu ve mesaj trafiğini önemli ölçüde azalttığı da görülmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . iii
ABSTRACT . . . iv
ÖZET . . . vi
LIST OF FIGURES . . . x
LIST OF TABLES . . . xii
LIST OF SYMBOLS/ABBREVIATIONS . . . xiii
1. INTRODUCTION . . . 1
2. OVERVIEW OF WIRELESS OVERLAY NETWORKS . . . 3
2.1. Overlay Networks . . . 3
2.2. Wireless Overlay Networks . . . 3
2.3. Main Challenges and Problems . . . 5
2.4. Horizontal and Vertical Handovers . . . 7
3. OVERVIEW OF MOBILE IPV6 . . . 11
3.1. Components of MIPv6 . . . 11
3.1.1. Nodes . . . 11
3.1.2. Links . . . 11
3.1.3. Addresses . . . 11
3.2. Communication Between MIPv6 Components . . . 12
4. PRESENT HANDOVER SCHEMES . . . 14
4.1. Basic Horizontal Handover Scheme . . . 14
4.2. Basic Vertical Handover Scheme . . . 15
4.3. Related Work . . . 17
4.3.1. Mobility Management Architectures . . . 17
4.3.2. Handover Decision Algorithms . . . 18
4.3.3. Testbeds . . . 20
5. A NEW MOBILITY MANAGEMENT SCHEME FOR WIRELESS OVERLAY NETWORKS . . . 22
5.1. Horizontal Handover Scheme . . . 22
5.1.1. Proposed Solution . . . 22

5.1.2. Dealing With Oscillations 23
5.1.3. Comparison with Anticipated Handover 24
5.1.4. Simulation and Results 24
5.2. Vertical Handover Scheme 26
5.2.1. Proposed Architecture for WON Interconnection 26
5.2.2. Vertical Handover Phases 28
5.2.3. Performance Evaluation 35
6. CONCLUSION AND FUTURE WORK 48
REFERENCES 50

LIST OF FIGURES

Figure 2.1. Hierarchical structure of wireless overlay networks 5

Figure 2.2. Roaming in wireless overlay networks 8

Figure 2.3. Horizontal handover 10

Figure 2.4. Upward vertical handover 10

Figure 3.1. Communication of MIPv6 components 13

Figure 4.1. Present scheme for horizontal handover 14

Figure 4.2. Overview of the basic handover system 15

Figure 5.1. Proposed scheme for horizontal handover 23

Figure 5.2. Fluctuating signal strengths causing a ping-pong effect 24

Figure 5.3. Horizontal handover latency vs. number of mobile nodes 25

Figure 5.4. WLAN-WWAN interconnection architecture based on IP 27

Figure 5.5. Vertical handover decision algorithm 28

Figure 5.6. Downward vertical handover signaling flow 31

Figure 5.7. Upward vertical handover signaling flow 33

Figure 5.8. Map of the simulation environment 35

Figure 5.9. Average number of handovers an MN performs v.s. MN's speed . 38

Figure 5.10. Cost of handover decision calculation 39

Figure 5.11. MNs' average throughput v.s. MNs' speed 40

Figure 5.12. Badis' hierarchical mobility management 41

Figure 5.13. Badis' handover signaling flow 42

Figure 5.14. Wang's integrated UMTS-WLAN architecture 43

Figure 5.15. Handover from UMTS to WLAN in Wang's architecture 44

LIST OF TABLES

Table 2.1. Characteristics in existing and emerging wireless technologies . . . 4

Table 5.1. Message lengths for horizontal handover 23

Table 5.2. Message lengths for vertical handover 33

Table 5.3. Simulation parameter values for access providers 36

Table 5.4. Simulation parameter values for HDM algorithm 37

Table 5.5. Cost of operations 40

Table 5.6. Length of messages in Badis's architecture 42

Table 5.7. Length of messages in Wang's architecture 43

Table 5.8. Approximate latency of messages in each architecture 46

LIST OF SYMBOLS/ABBREVIATIONS

bitRate_j	Bitrate of wireless network j
BW	bandwidth
Cost_j	Cost of wireless network j
Coverage_j	Coverage of wireless network j
L	latency
L_{Badis}	upward vertical handover latency in Badiss architecture
L_L	latency encountered in sending a 128-bit message in lower overlay
L_{NI}	latency of network interface
$L_{release}$	connection release time
L_{SA}	upward vertical handover latency in SA-based architecture
L_{setup}	connection setup time
L_U	latency encountered in sending a 128-bit message in upper overlay
L_W	latency encountered in sending a 128-bit message in wired network
L_{Wang1}	upward vertical handover latency in Wangs architecture (same administrative domain)
L_{Wang2}	upward vertical handover latency in Wangs architecture (different administrative domains)
Load_j	Traffic load of wireless network j
n	path loss exponent
n_{buf}	number of buffered packets sent from new wireless network to MN
PL(d)	path loss at distance d
P_t	transmitted power
RSS(d)	received signal strength at distance d
RSSI_j	wireless network j's received signal strength at the mobile node
S	path loss constant depending on the propagation environment

S_{MSG}	size of message in bits
$speed_{MN}$	Mobile Node's speed
$T_{RSSHIGH}$	RSS at distance (0.5 × coverage)
T_{RSSLOW}	RSS at distance (0.8 × coverage)
T_{RSSUP}	RSS at distance (0.8 × coverage)
$T_{timerDOWN}$	threshold timer value for monitoring the consistent utility improvement in the target wireless network
$T_{timerUP}$	threshold timer value for monitoring the consistent RSS decay
$Utility_j$	Utility of wireless network j
w_B	weight of $bitRate_j$
w_C	weight of $cost_j$
w_L	weight of $Load_j$
w_{RSSI}	weight of $RSSI_j$
X_σ	zero-mean Gaussian random variable with standard deviation σ

AP	Access Point
AR	Access Router
BM	Billboard Manager
BS	Base Station
CN	Correspondent Node
DAD	Duplicate Address Detection
FA	Foreign Agent
GPRS	General Packet Radio Service
GSM	Global System Mobile
HA	Home Agent
HAP	High Altitude Platform
HDM	Handover Decision Manager
IP	Internet Protocol
LOSA	Lower Overlay Subnet Agent
MAC	Media Access Control
MIP	Mobile Internet Protocol

MN	Mobile Node
NAV	Network Allocation Vector
NI	Network Interface
QoS	Quality of Service
RA	Router Advertisement
RF	Radio Frequency
RSS	Received Signal Strength
SA	Subnet Agent
SNR	Signal-to-Noise Ratio
TCP	Transmission Control Protocol
TFRC	TCP-Friendly Rate Control
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications Systems
UOSA	Upper Overlay Subnet Agent
VPN	Virtual Private Network
WLAN	Wireless Local Area Network
WWAN	Wireless Wide Area Network
WON	Wireless Overlay Network

1. INTRODUCTION

Wireless networking is becoming an increasingly important and popular way to provide global information access to users regardless of time and location. However, there is no unified wireless technology which fits all user requirements at all times. Instead, using several overlaying wireless networks can provide the best possible data delivery service. Therefore, extensions to a traditional cellular handover (handoff) system are made in order to handle the simultaneous operation of multiple wireless network interfaces. This new system allows mobile users to roam in a *Wireless Overlay Network (WON)* structure which consists of several network layers. In such a structure, the lower layers contain high-bandwidth wireless cells covering a small area whereas higher layers contain lower bandwidth wireless cells covering a larger area. The users can connect to the wired network through multiple wireless subnets, and this offers the best possible connectivity based on the users' geographic location and local wireless connectivity.

In spite of the WON's advantage of global connectivity, users may experience some disturbance in service during a *horizontal handover* between cells of the same wireless network or a *vertical handover* between different overlay networks. Data packets may be lost during these handover periods. Therefore, constructing an efficient mobility management system and making the handovers as seamless as possible is crucial for critical wireless applications.

This thesis covers the problem of building an efficient mobility management system which reduces handover latency, handover overhead and the number of unnecessary handovers while considering power and throughput constraints. Our ambition is to achieve seamless handovers in a WON without requiring the end users to make additional configurations. The first goal of this thesis is to study the current handover schemes and some enhancement ideas that may be applied to these schemes. It is also intended to summarize prior work done in this area. The second goal of the thesis is to present a new mobility management scheme to provide a seamless service to roaming

end-users. Finally, the thesis also aims to present a simulation model evaluating the performance of the proposed schemes and comparing it with the current schemes.

Chapter 2 gives an overview of WONs and describes the main challenges they pose as well as horizontal and vertical handovers occuring within them. Chapter 3 gives an overview of Mobile IPv6 by discussing the MIPv6 components and how they communicate. This chapter is essential in comprehending the related and proposed work which uses MIPv6 components. Chapter 4 presents the currently used basic horizontal and vertical handover schemes as well as their drawbacks. The chapter also summarizes prior work done for WONs. Chapter 5 presents the new mobility management scheme proposed for WONs. This scheme considers both horizontal and vertical handovers and proposes new solutions for reducing power usage, handover latency and handover overhead as well as eliminating the number of unnecessary handovers. This chapter also evaluates the performance of the proposed scheme by various simulations and analytic work. Finally, Chapter 6 concludes the thesis and discusses future work.

2. OVERVIEW OF WIRELESS OVERLAY NETWORKS

2.1. Overlay Networks

An overlay network is a virtual network which is built on top of an existing network in order to add new functionality or network service not available in the existing layers [1]. The Internet itself is an overlay network which is built on top of local area networks (e.g. Ethernet) and telephone networks, and this new overlay adds an Internet protocol header to all the packets. Overlays are beneficial in deploying new routing and packet management algorithms quickly and easily instead of upgrading the existing layers which is very expensive. Another benefit of overlay networks is that any bug in the overlay network is usually isolated from the lower overlays.

As stated in [1], there are mainly two types of overlays: i) Routing Overlays and ii) Structured Overlays. Routing overlays intend to modify and improve the routing in the existing overlays so that new functionality and services may be added. Virtual Private Networks (VPN) can be considered as a routing overlay achieving better security and authentication. Structured overlays are used for combining large distributed systems. These overlays often replicate content across many individual nodes so that the network does not collapse due to the failure of individual links and nodes.

2.2. Wireless Overlay Networks

Wireless Overlay Networks (WONs), which were first introduced in [2], aim to combine different wireless technologies in order to provide wide-area coverage, best possible bandwidth and lowest possible latency to all mobile users. WONs are constructed when different wireless network technologies developed for different purposes are overlaid to form larger wireless networks. Table 2.1 depicts a list of some different wireless technologies [3, 4]. The need to unify all these technologies so that mobile users can seamlessly roam between access networks brings up the idea of heterogenous networks and WONs.

Table 2.1. Characteristics in existing and emerging wireless technologies

Network	Coverage	Data Rates	Cost
Satellite	World	Max. 144 kb/s	High
GSM/GPRS	Approx. 35 km	9.6 kb/s up to 144 kb/s	High
IEEE 802.16a	Approx. 30 km	Max. 70 Mb/s	Medium
IEEE 802.20	Approx. 20 km	1-9 Mb/s	High
UMTS	20 km	up to 2 Mb/s	High
HIPERLAN 2	70 up to 300 m	25 Mb/s	Low
IEEE 802.11a	50 up to 300 m	54 Mb/s	Low
IEEE 802.11b	50 up to 300 m	11 Mb/s	Low
Bluetooth	10 m	Max. 700 kb/s	Low

WON has a hierarchical structure with different levels. These levels differ from each other by their coverage areas, bandwidths and latency. Higher levels have a larger coverage area but lower bandwidth. Conversely, lower levels contain high-bandwidth cells covering a small area. Overlaying a high bandwidth in-room infrared network with a radio frequency network having less bandwidth and thus establishing connectivity between rooms is a WON example mentioned in [2]. From Table 2.1, it seems that Satellite networks cover WLANs. However, it should be considered that Satellite networks are not always available especially indoors.

It is desirable that wireless applications operate in lower levels of a WON since they provide greater bandwidth. However, providing a high bandwidth to all users all the time is not possible. Therefore, low-mobility users are served by lower layers and benefit from a greater bandwidth whereas high mobility users are served by upper layers and benefit from a greater coverage area.

For example, IEEE 802.11 and HIPERLAN/2 are two wireless local area network (WLAN) standards whereas GSM/GPRS and UMTS are designed for wide area communication. These technologies vary in frequency band, bandwidth, data transmission latency, coverage range, etc. A WON structure covering all these technologies contains IEEE 802.11 and HIPERLAN/2 at lower levels, and GSM/GPRS and UMTS at higher

levels.

Fig. 2.1 is an example of a general WON structure [2]. In this figure, the regional area has a larger coverage area but a lower bandwidth than all the lower layers.

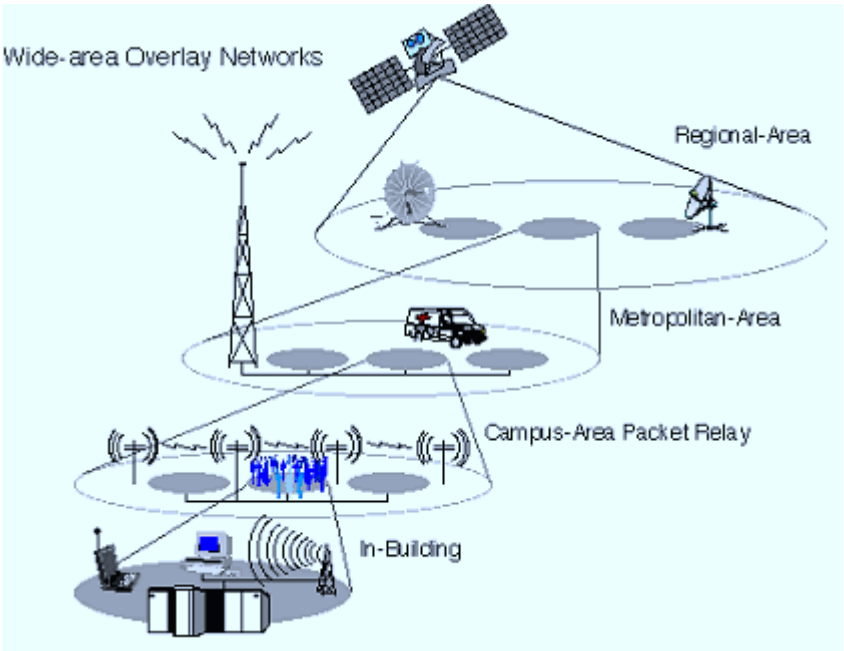


Figure 2.1. Hierarchical structure of wireless overlay networks

In a WON structure, the MN travels within or between the different overlays. Therefore, it must have a switching ability as well as the necessary network interfaces which are used interchangeably after switching between different networks.

2.3. Main Challenges and Problems

Since a WON is composed of a variety of wireless technologies with different coverage, bandwidth and latency, it poses many challenges. Some of these challenges stated in [5, 6, 7] are as follows:

- Mobile users access network through a wireless link. Thus, they experience a lower bandwidth and higher error rate when compared to wired users. Besides, mobile users may not be in the coverage area of any network sometimes so the network may be unavailable in such situations. Therefore, a wireless network

may be viewed as unreliable in terms of availability.

- High handover latency impairs both applications and protocols transfering information. For example, various testbed and simulation measurements show that TCP and TFRC (TCP-Friendly Rate Control) performance is significantly reduced during vertical handovers [4]. Sessions must be seamlessly maintained as the MNs roam among different overlays. That is, the handovers must be fast enough not to cause any significant service degradation. The main goal of the wireless overlay networking is to allow a user to use fully-interactive multimedia communication tools across all of these different network interfaces eventhough the networks provide different levels of service. For example, the infrared network may support full-motion video and high-quality audio, the RF network may support a lower frame-rate video and lower quality audio, and the wide-area network may support only audio. Therefore, making the switch between networks as seamless as possible for disruption intolerant applications and with as little data loss as possible means that achieving "low latency handover" is really a difficult issue.
- In order to achieve a truly ubiquitous and seamless access to information systems, end users should not deal with additional configurations.
- The simplest approach to managing multiple wireless network interfaces is to keep all of them on all the time. Measurements of commercially available wireless network interfaces [8] show that keeping an IBM Infrared and WaveLan RF interface on all the time consumes approximately 1.5 watts. Therefore, another challenge is "power saving" in the case of vertical handovers. The aim is to minimize the power drain due to simultaneously active multiple network interfaces because mobile devices have limited processing and power capabilities when compared to desktop computers.
- There is no comparable signal strength available to aid the decision of the "best" network to make vertical handover, because the networks have such varying characteristics. For example, an in-building RF network with a low signal strength may still yield better performance than a wide-area data network with a high signal strength [8]. Thus, discovering the right time to perform handovers in a wireless channel is difficult to predict and characterize.

- During a handover procedure, upper-layer applications are interested in network conditions such as available bandwidth, delay and user preferences rather than the physical layer parameters such as received signal strength and signal-to-interference ratio.
- Designing vertical handovers in wireless overlays requires bandwidth overheads in the form of beacon packets and handover messages that are necessary to provide service to roaming users. The cost of this additional network traffic should be minimized while also providing minimal disruption for transitions between networks.
- Media Access Control (MAC) protocols, which are significant for the performance and use of wireless networks, are different for the cellular systems in U.S. and Europe. These differences affect the interoperability of different networks and hence global roaming of mobile users.

To sum up, an efficient mobility management system in WONs must establish balance among handover latency, power consumption of the MN's network interfaces and the wasted bandwidth due to additional network traffic occuring during handovers [9].

2.4. Horizontal and Vertical Handovers

Two types of handovers occur as the MN travels within or between the different overlays of a WON: horizontal and vertical. Both handover types that occur as a user roams in a WON are illustrated in Fig. 2.2 [2].

Horizontal (or standard) handovers are the conventional handovers which take place while the MN switches between cells of the same network overlay. Therefore, the MN does not need to change the network interface it is using. Horizontal handovers are mandatory since a MN, which moves very far away from its base station (BS), cannot continue its communication without performing a horizontal handover.

Fig. 2.3 depicts a horizontal handover [9]. In this figure, there is a handover

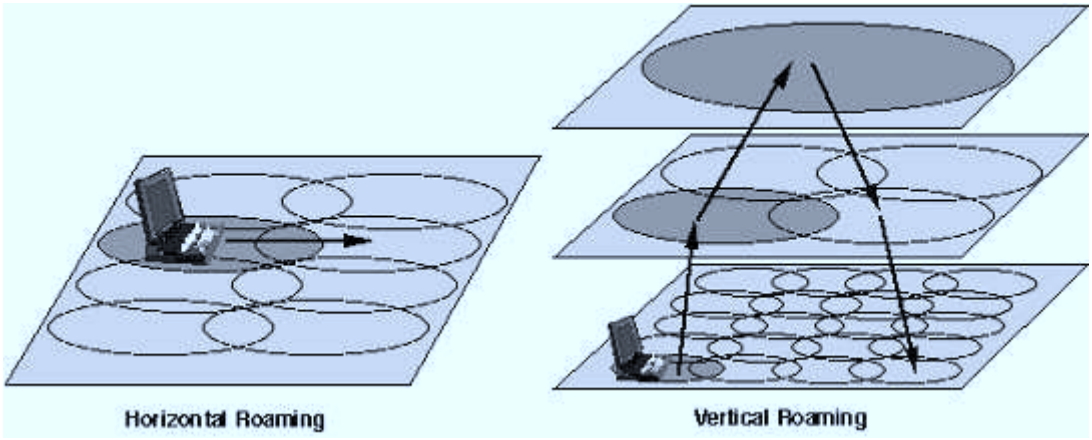


Figure 2.2. Roaming in wireless overlay networks

from the MN's old BS to its new BS; both of them having the same wireless network interface. The beacons, which are generated by the BSs, help the MN to determine whether a horizontal handover is necessary or not. The MN continously checks the signal strengths of all beacons. If a new BS sends signals stronger than the old one, the handover mechanism is triggered by the MN. The MN sends a handover request (or a greeting packet) to the new BS and receives an acknowledgement. The new BS also sends a notification to the old BS and waits for an acknowledgement from the old BS. Meanwhile, the old BS buffers the packets coming to the MN so that these packets, which arrive during handover, are not lost. The buffered data packets are sent to the new BS after the notification acknowledgement. As it is seen from Fig. 2.3, the handover latency is the time between the moment when the MN instructs the new BS for a handover start and the moment when the new BS sends the first data packet to the MN.

Unlike horizontal handovers, vertical (or inter-system/inter-network) handovers are not strictly mandatory. Instead, they occur as a result of the tradeoff between coverage and bandwidth. A vertical handover takes place when the MN moves from a cell in one overlay to a cell in a different overlay in WON. Moving from an IEEE 802.11 WLAN to a GSM cell is a typical example for vertical handover. After vertical handovers, the MN's network interface changes because it travels between different wireless communication architectures [10].

Vertical handovers are either upwards or downwards and start when the lower overlay becomes reachable or unreachable. An upward vertical handover occurs when the MN moves to an upper WON layer with larger coverage area but lower bandwidth. This happens as the MN becomes highly mobile and the lower layer becomes unreachable. On the other hand, a downward vertical handover occurs when the MN moves to a lower WON layer with greater bandwidth but smaller coverage area.

Fig. 2.4 shows how an upward vertical handover occurs [9]. It starts when the MN senses that the current layer is not reachable anymore. The MN sends "start forwarding" handover message to the upper layer, New Overlay. Then, it sends "stop forwarding" handover message to the current layer, Old Overlay, via the New Overlay. Handover ends when the first packet is received from the New Overlay. As depicted in Fig. 2.4, the handover latency is the time between the moment when the MN detects that the Old Overlay is unreacheable and the moment when the New Overlay starts forwarding data packets to the MN.

Generally, vertical handovers are analysed in two main steps: i) a *handover decision process* and ii) a *handover execution process* [11]. In the handover decision process, the time to perform the handover is decided by the MN, network or both. In the handover execution process, the handover takes place. Still, there are some approaches which consider an additional first step called *system discovery* in which the MN discovers the reachable wireless networks [12].

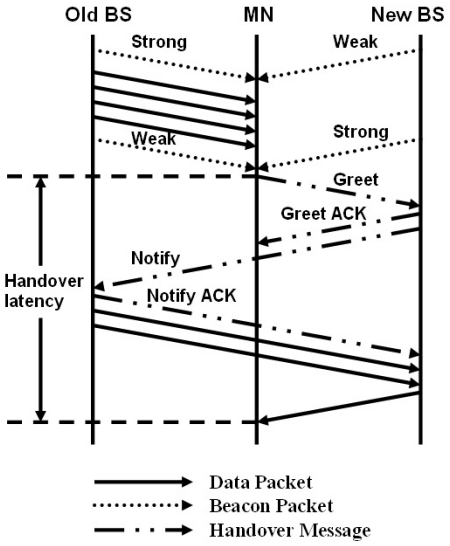


Figure 2.3. Horizontal handover

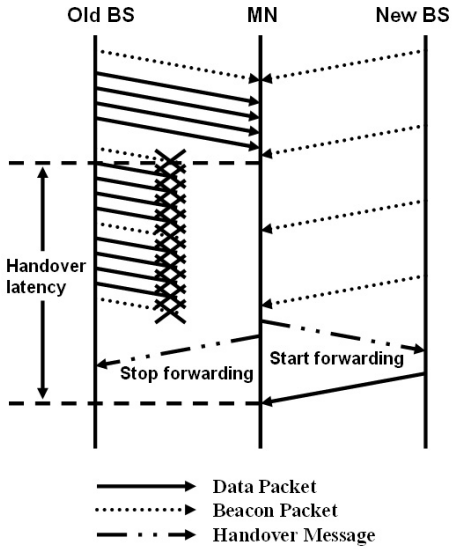


Figure 2.4. Upward vertical handover

3. OVERVIEW OF MOBILE IPV6

A MIPv6 node can continuously change its location on an IPv6 network and still be connected. The connectivity of the node is maintained by various messages going back and forth between the MIPv6 components [13, 14]. This section discusses the MIPv6 components and how they communicate.

3.1. Components of MIPv6

3.1.1. Nodes

The *Mobile Node* (MN) is the node that travels between IPv6 subnets. It acquires new temporary addresses as it roams into new subnets and maintains its connectivity by using its unchanging home address. The *Correspondent Node* (CN) is the node that wants to exchange data with the MN. The *Home Agent* (HA) is the router on the MN's home subnet and provides communication between MN and CN. If the MN is away from its home, it registers its current temporary address with the HA and the HA forwards data, which comes to the MN's home address, to the MN's temporary address. The *Foreign Agent* (FA) is the home agent of the foreign subnet the MN roams into. HA and FA are both Access Routers (AR) or Access Points (AP).

3.1.2. Links

Home Link is established between MN and HA. The MN obtains its home address from this link. *Foreign Link* is between MN and FA. The MN obtains its temporary address from this link.

3.1.3. Addresses

The 128-bit IPv6 addresses consist of a 64-bit routing prefix and a 64-bit interface identifier. The routing prefix is used for routing the packet to the correct subnet

whereas the interface identifier identifies the specific node within the network. *Home Address* always identifies the MN even if it is outside its home subnet. It is used on Home Link. On the other hand, *Care-of Address* is the temporary address assigned to the MN when it is attached to a foreign link and provides information about the MN's current location. It is used on Foreign Link.

3.2. Communication Between MIPv6 Components

Fig. 3.1 illustrates the communication lines between the components of MIPv6 [13]. In this figure, the MN travels from its home subnet to a foreign subnet, registers with the foreign subnet and gets a new address called the care-of address (1). Then, the MN transmits this address to the HA (4) and sometimes to the CN (5) if the CN is MIPv6 capable. This message, which enables a mapping between the MN's home address and care-of address, is called a *binding update*. Then, HA sends a binding acknowledgment (2) to the MN. If the CN is not MIPv6 capable, it sends data to the MN's home-address and HA collects the data to deliver it to the MN's care-of address (3). The MN sends its response to the CN over HA but this type of communication increases the burden of routing at HA. On the other hand, if the CN is MIPv6 capable, it may directly send data to and receive data from the MN's care-of address (5). As a result of binding, it can be said that there is a dynamic tunnel between home address and care-of address at any time.

Moreover, it is worthy to note that the FA periodically sends a message called the Router Advertisement (RA) message. This contains the globally unique routing prefix that is used to formulate a care-of address by the MN. As soon as the MN switches its network, it can send a Router Solicitation message to the FA before receiving the periodic Router Advertisement message.

In terms of MIPv6, handover latency can be divided into three main components: i) *detection period*, ii) *address configuration period* and iii) *network registration time* [11, 7]. During the detection period, the MN discovers new wireless networks due to the new RAs. During the address configuration period, the MN assigns its interface

a new care-of address based on the prefix of the new wireless network. Finally, the network registration time is the time between sending a binding update to the HA as well as the CN and receiving the first packet from the CN via the new network.

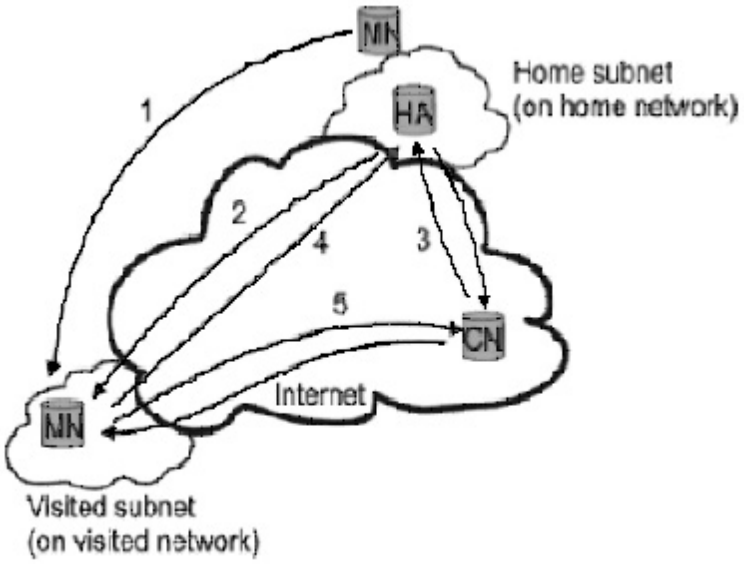


Figure 3.1. Communication of MIPv6 components

4. PRESENT HANDOVER SCHEMES

4.1. Basic Horizontal Handover Scheme

Previously, it has been explained that MN registers with the foreign subnet and gets a new care-of address as soon as it enters a new foreign subnet. Fig. 4.1 shows this registration process. As explained in [15], first, MN creates its own care-of address by making use of RA messages sent from FA. Then, MN sends a request to FA to check the uniqueness of the care-of address it has just created (1). FA performs Duplicate Address Detection (DAD), which takes a considerable amount of time, and sends the result back to MN (2). Finally, if everything is OK, i.e. if the newly created care-of address is unique, the MN registers with the foreign subnet (3 and 4). Otherwise, MN has to create another care-of address and the same procedure starts all over again.

One of the major drawbacks of this model is certainly starting the handover process after the MN enters the new subnet. No matter how fast the signals are transmitted or how fast the DAD operation is done, it is impossible to get rid of the handover latency. If the handover process could start before the MN actually enters the new subnet, then the handover latency can be decreased to some extent.

Secondly, DAD takes a very long time and impairs the handover performance. Currently, there are some studies towards analysing the origin of this performance

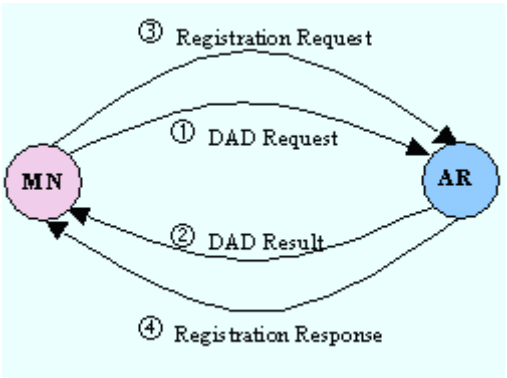


Figure 4.1. Present scheme for horizontal handover

degradation and some proposed ideas to finish DAD earlier [16, 17]. If the AR could keep track of all the care-of addresses in its coverage area and could generate a new care-of address for the MN, then time would not be wasted for the DAD operation.

The basic horizontal handover scheme may be enhanced in such a way that it adopts MIPv6 and consists of soft handover mechanisms not imposing any change to the MIPv6 standard [18]. For this solution, a new component called "Duplication & Merging Agent" (D&M), which is a router located at the core network to duplicate and merge IPv6 flows, is introduced. At first, packets going from the CN to the MN are duplicated and tunnelled to the MN via corresponding ARs and then these duplicates are filtered at the MN. By receiving data from multiple ARs simultaneously, the MN can move from one cell to another without significant interruption. However, such a solution cannot prevent the long handover latency due to the DAD operation.

4.2. Basic Vertical Handover Scheme

In [8], a model of handovers which is built on top of the mobile routing capabilities of MIP is proposed. This model is based on the infrastructure that is described in [19] and the MIP specification that is described in [20]. The architecture of the basic handover scheme is shown in Fig. 4.2.

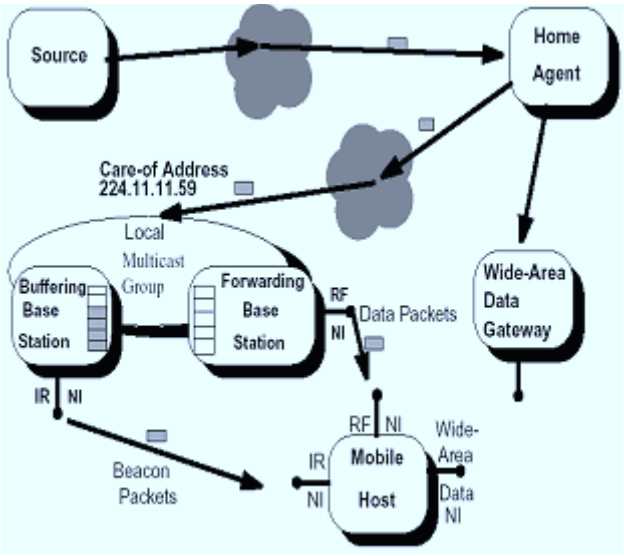


Figure 4.2. Overview of the basic handover system

In this structure, MNs connect to a wired infrastructure via Base Stations (BSs). An HA performs the same functions as in Mobile IP. The difference is that the care-of address is not a unicast address but an encapsulating multicast address. The MN is responsible for initiating handovers between BSs and between different network interfaces. A small group of BSs are selected by the MN to listen on this multicast address for packets encapsulated and sent by the HA. One of the BSs is selected by the MN to be a forwarding BS; it decapsulates and forwards the packets it receives on the multicast address to the MN. The other BSs are buffering BSs; they hold a small number of packets from the HA in a circular buffer. When the MN initiates a handover, it informs the old BS to switch from forwarding to buffering mode. The new BS then forwards the buffered packets that the MN has not received yet. For networks where the BS infrastructure is not under control, the HA acts as the BS for the MN; the HA sends separate unicast packets to the care-of address of the MN's wide-area data interface [8].

The BSs send out periodic beacons similar to Mobile IP foreign agent advertisements. The MN listens to these packets and decides which BS should be forwarding packets and which BSs should be buffering packets in anticipation of a handover, and which BSs should be members of the multicast group assigned for a single MN. The signal strengths of these beacon packets are compared and the BS with the highest signal strength is chosen as the forwarding BS.

Upward vertical handovers are initiated when several beacons on the currently connected network are not received. Then, the MN decides that the current network is not reachable and is handed over to the upper overlay network. Downward vertical handovers are initiated when several beacons in a row are heard from a lower overlay's network interface (NI). The MN decides that the MN is now within the range of the lower overlay's NI and switches to the lower overlay [8].

This basic handover scheme may be enhanced by several ways. For example, in order to prevent the mobile users from experiencing a significant drop in QoS, priority-based algorithms may be deployed as stated in [21]. By this way, mobile users running

latency-sensitive applications (e.g. transferring digitized voice, video, etc.) have the priority to be served first and do not experience a significant performance degradation.

Secondly, policy-based call admission controls may be deployed in order to decide when the MN should perform a vertical handover [22, 23]. Such a policy-based scheme allows vertical handover requests and admits MNs into the requested overlays based on required QoS level, residual capacity in each available network, bandwidth, coverage and cost. Policy Decision Points (PDPs) implemented in base stations determine the network health by monitoring video frame drop rate (FDR), block error rate (BLER) and voice call blocking. If the defined policies find the network health to be convenient, then the MN performs a vertical handover. Such a solution prevents unnecessary handovers and achieves a balanced load between the available access networks.

4.3. Related Work

Wireless Overlay Networks (WONs) were first introduced by Brewer and Katz [2]. Related work in this area focuses on the development of new mobility management architectures as well as improving the handover decision phase to quickly and accurately trigger handovers without triggering any unnecessary ones.

4.3.1. Mobility Management Architectures

The idea of vertical handovers were first introduced by Stemm and Katz [8]. The vertical handover approach proposed by Seshan et al. is very similar to that of Stemm and Katz [24]. In both approaches, the MN selects one of the BSs to be a forwarding BS and the other BSs to be buffering BSs. The buffering BSs buffer the recent few packets and forward them to the mobile node during a handover.

Badis and Agha proposed a hierarchical mobility management scheme which uses multicasting as a packet forwarding mechanism [9, 25]. Such a multicasting solution handles frequent handovers but increases wasted bandwidth. Besides, the scheme proposes two new classes of MNs: i) high-mobility and ii) low-mobility. An MN which

exceeds a certain *speed threshold* connects to the next higher overlay if there are available resources. Such an approach decreases the number of handovers.

Buddhikot et al. considered only IEEE 802.11 and 3G wireless networks and described two approaches for the integration of these two technologies [26]. In the *tight integration approach*, the 802.11 network appears to the 3G core network as another 3G access network. Thus, the 802.11 network emulates functions natively available in 3G. Besides, the MNs need to implement the 3G protocol stack on top of their 802.11 network cards. In the *loose integration approach*, which is shown to be better, 802.11 and 3G networks can be independently deployed by the introduction of the IOTA (Integration Of Two Access technologies) 802.11 gateway element. This gateway serves multiple 802.11 access points in a *hot-spot* and helps the MN to obtain 802.11 access as it roams into this hot-spot.

Wang et al. proposed an architecture based on MIPv6 for integrating UMTS and IEEE 802.11 WLAN [27]. In this architecture, UMTS and WLAN are connected to each other by routers in the same administrative domain without any master-slave relationship. The routers connecting these different networks act as the mobility gateways responsible for the mobility management of the whole network.

4.3.2. Handover Decision Algorithms

Related work on handover decision suggests different techniques. Traditional decision algorithms based on received signal strength (RSS), thresholds and dwell timers are some examples [28].

Pahlavan et al. considers a 2-layer WON and suggests to measure the relative bandwidths of WLAN and GPRS and uses neural networks in determining the handover decision time [28]. However, no information on how to train and use the neural network is given.

Zhang et al. again considers a 2-layer WON consisting of IEEE 802.11 WLAN

and GPRS WWAN, and uses different methods for switching to WLAN and WWAN [29]. When switching to WWAN, the decision algorithm periodically monitors RSS and tries to detect a decay in RSS. This lets the MN to switch to WWAN very quickly as the WLAN becomes unavailable. When switching to WLAN, the algorithm uses the network allocation vector (NAV) in the MAC layer to estimate network conditions (e.g. available bandwidth and access delay) because the NAV busy state can reflect the media's busy state or traffic load to some extent. This lets the MN to decide handover by being aware of network conditions. Although the algorithm is smart and simple, it is not suitable for different types of wireless networks which do not use NAV. Besides, the algorithm does not consider many factors such as the network's coverage, cost, mobility support, etc.

Murray et al. introduces fuzzy logic into handover decision mechanism and proposes policies which are deployed on the MN as well as the stationary policy decision points in the network [23]. In this approach, the decision is made by calculating the network's health from video frame drop rate, block error rate and voice call blocking rate. However, the proposed solution does not consider the network's coverage, cost, mobility support, power consumption, etc. Moreover, the deployment of policy decision points in the WON increases the complexity of the mobility management scheme considerably.

Some approaches use signal-to-noise ratios (SNR) together with conventional two-level thresholding scheme [26, 30, 31]. This scheme lets the MN to stay with the current wireless network as long as possible.

Chen et al. presents an adaptive scheme which makes use of utility functions [12]. As depicted in Equation 4.1, the utility function of wireless network j is composed of several normalized factors $f_{i,j}$ multiplied by their weight or importance w_i.

$$Utility_j = \sum_i w_i \times f_{i,j} \tag{4.1}$$

The handover decision algorithm calculates and compares the utilities of candidate networks with the current network for a certain amount of time called the stability period. The stability period changes according to handover latency and the utility ratios. Thus, if the utility of the candidate network is increasing relatively more, then the stability period decreases and the handover decision is made earlier. Otherwise, the stability period increases and the MN avoids unnecessary handovers. In spite of its various advantages, this scheme, which continuously calculates complex utility functions during the stability period, consumes more MN power relative to the other methods and neglects the computational advantage of performing different tasks for upward and downward vertical handovers. Besides, the algorithm needs to estimate the handover latency in order to calculate the stability period. Estimating this latency before even doing a handover is unreliable and may lead to erroneous results.

Finally, there is an ongoing research in integrating fuzzy logic to handover decision algorithms [32, 33, 34]. These systems take several parameters such as current and previous RSS, MN speed, network traffic parameters as input. Then, according to the fuzzy rules defined in a series of fuzzy logic systems, several output parameters such as low and high threshold RSS values are generated dynamically. Finally, these output parameters are used in making a handover decision.

4.3.3. Testbeds

There are also some research studies evaluating the horizontal and vertical host mobility using real testbed experimentation. Among these, Berkeley's Bay Area Research Wireless Access Network (BARWAN) project implemented in San Francisco Bay Area is the first [35]. This testbed consists of five overlays (in-room infrared, in-building radio frequency, campus/metropolitan area packet relay network, wide-area

packet switched data network and regional-area satellite data network). However, it relies on MIPv4. Moreover, it considers vertical handovers between Metricom Ricotech and WaveLAN [8]. Later on, some enhancements are implemented in this testbed in order to reduce handover latency. For example, the proposed handover protocol in [24] uses multicasting and intelligent buffering at the base stations to achieve a higher performance. However, these enhancements are based on PC base stations and IBM ThinkPad mobile hosts only. Currently, there are some studies of QoS enhancements on this testbed [36]. These studies try to enhance Internet QoS without any support from the underlying IP network. Smoothing losses, prioritization of packets and statistical loss and bandwidth guarantees are the three QoS enhancements studied currently.

Another testbed is the MIPv6 based LAN-WLAN-GPRS testbed implemented as part of the Cambridge Open Mobile Systems (COMS) project [37, 3]. It is called the LCE-CL testbed and built by the joint contribution of Laboratory for Communication Engineering (LCE) and the Computer Laboratory at the University of Cambridge. The current GPRS network infrastructure is produced by Vodafone UK whereas the WLAN access points are IEEE 802.11b capable. The testbed is used in evaluating the performance of vertical handovers between LAN $\leftrightarrow$ GPRS and WLAN $\leftrightarrow$ GPRS. The testbed does not have any special components or solutions to improve handover performance. However, some enhancement ideas are proposed such as increasing the frequency of Router Advertisements and Client-based Router Advertisement caching to decrease the MNs' detection time of WLANs [11].

5. A NEW MOBILITY MANAGEMENT SCHEME FOR WIRELESS OVERLAY NETWORKS

In order to minimize any disruption in connectivity, a proper mobility management scheme for WONs must take both horizontal and vertical handovers into account. Therefore, this section proposes novel solutions for decreasing handover latency and overhead both in horizontal and vertical handovers.

5.1. Horizontal Handover Scheme

5.1.1. Proposed Solution

The horizontal handover model proposed in this thesis is based on the one presented in [38] and focuses on the two major drawbacks of the present handover scheme described earlier. First, it reduces the handover latency significantly since the MN's handover process starts before the MN physically enters a new subnet. Moreover, since the care-of addresses are generated by the ARs instead of the MNs, the enormous amount of time wasted for DAD is also gotten rid of.

According to the proposed model depicted in Fig. 5.1, all the ARs know each other in advance. As the MN moves far away from its current AR (approximately 90% of the subnet's coverage radius), it sends a handover signal to its current AR noting that it is about to perform a handover (1). The handover signal also contains the MAC address of the MN and necessary information that identifies the target AR. Then, the current AR requests a care-of address for the MN from the target AR (2) by sending the MAC address of the MN. The target AR generates a new care-of address and sends it to the current AR (3). The target AR also reserves the generated address for a certain amount of time so that the address is not generated for other new comer MNs. The current AR informs the MN of its new care-of address (4). Having obtained its new care-of address, the MN registers to the new subnet (5, 6).

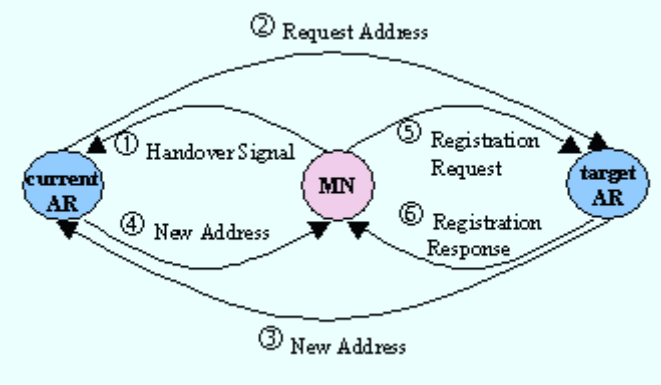


Figure 5.1. Proposed scheme for horizontal handover

Although this model decreases the horizontal handover latency considerably, it has some hidden costs. First, all the ARs have to be aware of each other in advance. Second, the messaging traffic is considerably increased since there is more signaling done in this model. Table 5.1 lists the messages used in horizontal handover.

Table 5.1. Message lengths for horizontal handover

Message	Length (bits)	Content
Handover Signal	256	MN's MAC address + target AR's IP
Request Address	128	MN's MAC address
New Address	128	MN's new care-of address
Reg. Request	128	MN's IP
Reg. Reply	128	target AR's IP

5.1.2. Dealing With Oscillations

Deciding the handover according to the relative signal strengths of two ARs is not always sufficient. If the MN moves back and forth between the two ARs, the received signal strengths from each AR fluctuates as shown in Fig. 5.2 and this causes multiple handovers back and forth. To prevent this oscillation called the *ping-pong effect*, the well known two-level thresholding scheme is used. According to this scheme, there are two threshold values, T_{LOW} and T_{HIGH}. The MN decides to handover if the received signal strength from the current AR falls below the threshold T_{LOW} or if the received signal strength from the target AR exceeds the threshold T_{HIGH}. Therefore, for the example depicted in Fig. 5.2, the MN decides to make a handover after point B where

the received signal strength from target AR exceeds the higher threshold, T_{HIGH}. By this way, the MN stays in the current AR as long as possible in order to prevent unnecessary handovers.

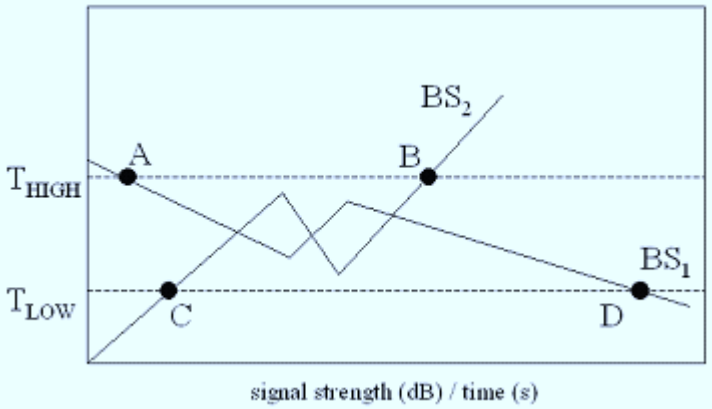


Figure 5.2. Fluctuating signal strengths causing a ping-pong effect

5.1.3. Comparison with Anticipated Handover

The proposed handover scheme is very similar to the *anticipated handover* described in [15]. However, in anticipated handover, no care-of address is requested from the target AR. Instead, the current AR generates a valid IPv6 address for the MN and sends it to the target AR for DAD operation and validation. Once the target AR checks the uniqueness of this address, it sends the validation result to the current AR. Therefore, anticipated handover does not eliminate the DAD operation at all.

5.1.4. Simulation and Results

In order to compare the performances of the two horizontal handover schemes explained earlier, a C++ simulation is implemented. In the simulation, the ARs are assumed to be base stations whose radius of coverage is approximately 500 meters. The base stations are assumed to support 50 MNs at most. The handover latencies are measured separately for each model as the MN moves from one base station to another. The simulation of the proposed model is performed for two different MN velocities: 20 m/s (approximate velocity of an automobile) and 2 m/s (approximate walking velocity).

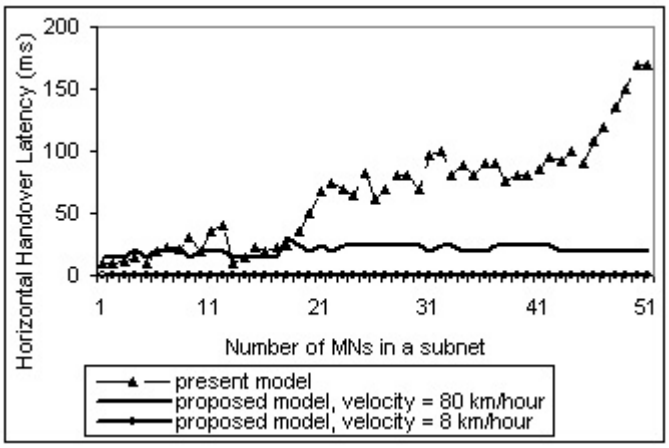


Figure 5.3. Horizontal handover latency vs. number of mobile nodes

Fig. 5.3 presents the performance of present solution and proposed solution for horizontal handover. Note that no value for MN's velocity is given for the present model. The handover latency of the present model is independent of the MN's velocity because the handover process starts after the MN enters the subnet. However, the MN's velocity plays an important role in the proposed model because the handover process starts as the MN's distance from the old AR approaches 90% of the AR's total coverage radius. Therefore, if the MN is not highly mobile, i.e. walking, it can be registered to the new subnet before even stepping into it and in such a case there is no handover latency.

As depicted in Fig. 5.3, the handover latency of the present solution increases as the number of nodes in the subnet increase. As the subnet approaches its limit (30-50 MNs), the probability of the MN's generating a unique care-of address decreases significantly. As a result, the MN tries to generate another care-of address and the DAD procedure starts all over again for that newly generated address.

However, in the proposed model, the time for generating a care-of address is independent of the number of MNs in a subnet. Therefore, the handover latency of the proposed model is almost constant.

Finally, note that if the number of MNs in a subnet is between 0 and 10, the present model's performance is better than that of the proposed model for MN's velocity

= 20 m/s. Therefore, one can conclude that the present model is more appropriate if the subnet does not contain many MNs whereas the model proposed in this thesis is more appropriate for highly crowded subnets.

5.2. Vertical Handover Scheme

In order to achieve a seamless vertical handover and maintain continuity of connection, a novel vertical handover management scheme is proposed in this section. This scheme integrates a *Handover Decision Manager* (HDM) to detect network condition changes in a timely and accurate manner, and a *Subnet Agent* (SA) architecture with a *Billboard Manager* (BM) that uses an end-to-end principle to maintain a connection without additional network infrastructure support. This section presents a completely IP-based approach to overcome the mentioned primary challenges of vertical handover to provide an efficient roaming service to end users.

5.2.1. Proposed Architecture for WON Interconnection

To interconnect WWAN and WLAN, an IP-centric architecture using MIPv6, as shown in Fig. 5.4 is proposed in this section. This architecture, which is a modification of the work presented in [39], requires each overlay to have a *Subnet Agent* (SA). For the sake of simplicity, Fig. 5.4 illustrates a 2-layer WON consisting of WLAN and WWAN only. In this figure, all Access Points (APs) of a specified region are linked to the SA of this level, that is, the *Lower Overlay Subnet Agent* (LOSA) and all Base Stations (BSs) of a wide-area region are connected to the SA of this overlay, that is, the *Upper Overlay Subnet Agent* (UOSA). These SAs are also connected to each other by fiber optic channels.

The SAs are newly designed for three purposes. First, they manage the handover process to reduce handover latency and to prevent data loss during roaming of the MN between different overlays. Second, they contain technical parameters such as access delay, available bandwidth, threshold values which help the newly introduced HDM to make an appropriate handover decision. Finally, they are responsible for maintaining

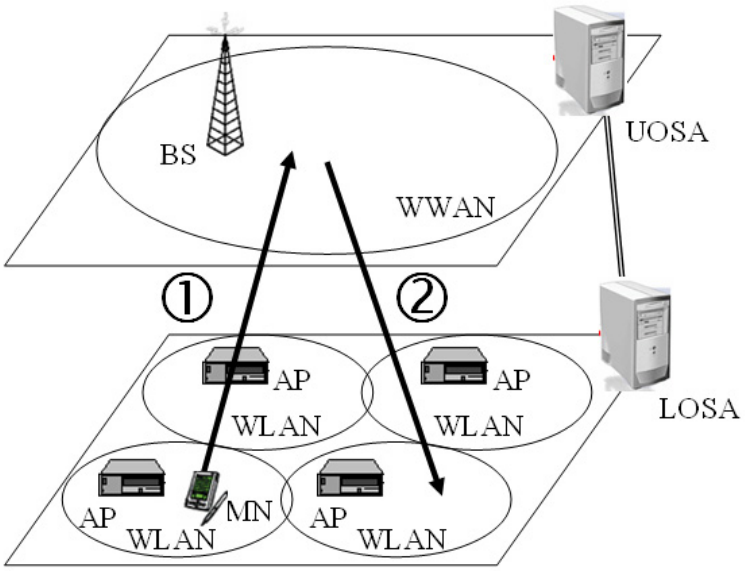


Figure 5.4. WLAN-WWAN interconnection architecture based on IP

the connection transparently and achieving the best possible communication quality.

There are two cases in which handovers occur in this architecture. First one is the upward vertical handover illustrated in case 1 of Fig. 5.4, in which the MN serviced in the WLAN moves to another area that is covered by WWAN. Second one is the downward vertical handover illustrated in case 2 of Fig. 5.4, in which the MN moves from WWAN to WLAN.

Dealing with Single Point of Failure. Providing only one SA for each overlay may lead to single point of failures. In order to make this architecture more resilient to such failures, the number of SAs responsible for an overlay may be increased. Such a solution requires a proper coordination between the SAs of the same overlay. Moreover, High Altitude Platforms (HAP), with a coverage radius of approximately 60 km, may be introduced to this architecture for backup. HAPs are balloons or small aircrafts situated in the stratosphere at around 20 km altitude. HAPs seem to become very popular in the near future especially in the delivery of broadband cellular communication systems [40]. Similarly, satellites may also be used for backing up SAs of higher overlays.

5.2.2. Vertical Handover Phases

Vertical Handover Decision Phase. Moving from an overlay to an underlay (e.g. WWAN to WLAN) is optional since the objective of the handover is to improve QoS rather than maintaining connectivity. Therefore, downward vertical handovers must be carried out cautiously by being aware of the network conditions. On the other hand, moving from an underlay to an overlay (e.g. WLAN to WWAN) is a result of the necessity to maintain connectivity. Therefore, we need a timely and accurate decision mechanism for upward vertical handovers.

In this thesis, a handover decision manager (HDM), which is inspired by the work in [29, 26, 12], is introduced to intelligently detect the conditions and availability of multiple networks. This HDM is also presented in [41]. Starting from lower overlays, HDM checks the availability of wireless networks and performs different tasks in order to decide different types of handovers. The flow chart of the handover decision algorithm proposed for HDM is depicted in Fig. 5.5.

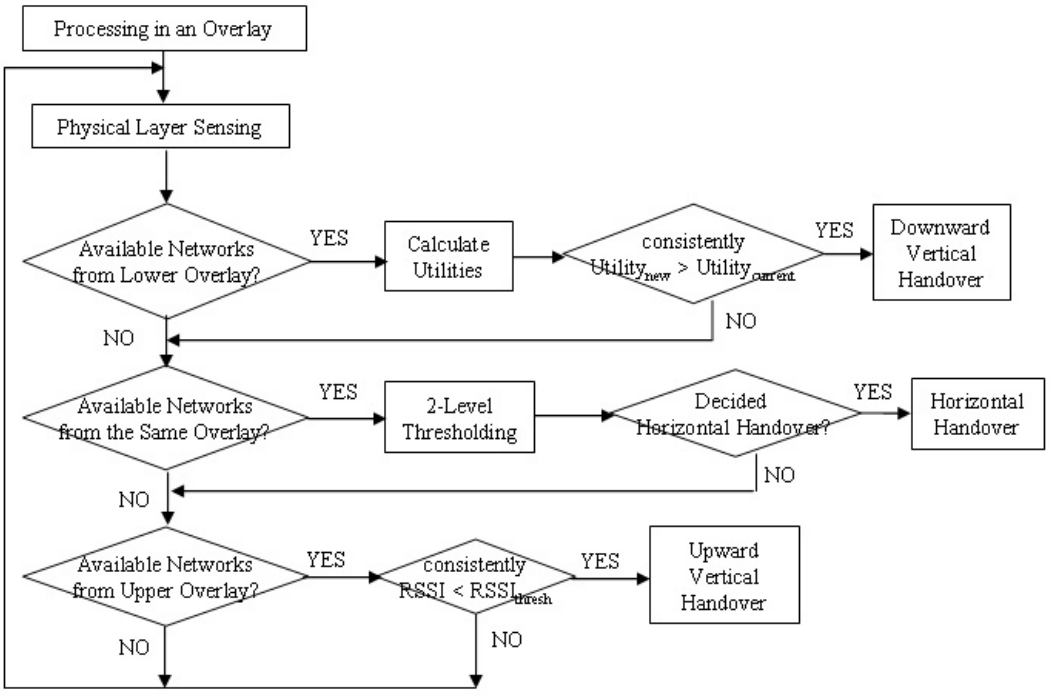


Figure 5.5. Vertical handover decision algorithm

Physical layer sensing is used to detect the availability of other wireless networks.

By dealing with networks from the lower overlays first, HDM gives priority to these networks.

First, HDM considers the availability of lower overlay networks in order to decide for downward vertical handover. If there are available lower overlay networks, it calculates the utilities of these candidates and compares these utilities with that of the currently serving network. The utility of a wireless network j, $Utility_j$, is calculated according to Equation 5.1.

$$\begin{aligned} Utility_j &= Coverage_j \times speed_{MN} + bitRate_j \times w_B + RSSI_j \times w_{RSSI} \\ &\quad + w_C/Cost_j + w_L/Load_j \end{aligned} \tag{5.1}$$

Equation 5.1 informs us about the network's coverage in kms ($Coverage_j$), bit rate in kbps ($bitRate_j$), normalized cost ($Cost_j$) and traffic load ($Load_j$) as well as the MN's received signal strength ($RSSI_j$). w_B, w_{RSSI}, w_C and w_L are the predefined weights for the factors $bitRate_j$, $RSSI_j$, $Cost_j$ and $Load_j$, respectively. These weights are determined either by the user or by the application running on the MN. On the other hand, the weight of the factor $Coverage_j$ is the MN's speed, $speed_{MN}$, and it is not defined by the application. As the MN's speed increases, it prefers to access wireless networks with more coverage. Therefore, the importance of coverage is directly proportional with the MN's speed. Here, the MN's speed is assumed to be determined by GPS.

If one of the candidates has a consistently better utility, then the HDM decides to make a downward vertical handover. In order to detect this consistency, the HDM facilitates from a timer mechanism. If the candidate network's utility is better than the current network's utility for a certain amount of time greater than or equal to a threshold timer value, then the HDM decides to handover to an underlay network. Here, the threshold timer value is determined by the application running on MN.

Secondly, HDM considers the availability of other wireless networks residing in the same overlay. It uses the conventional two-level thresholding scheme to decide

for horizontal handover. Since the horizontal handover decision mechanism is already explained, it will not be discussed here.

Finally, HDM considers the availability of upper overlay networks in order to decide for upward vertical handover. Since upward vertical handovers arise from the necessity of maintaining connectivity rather than improving QoS, we should quickly detect the unavailability of the currently serving network. That is, the aim of upward vertical handover is to switch to an upper layer before the current network link breaks. Therefore, calculating complicated utility functions and comparing them is unnecessarily time and power consuming.

In order to decide for upward vertical handovers in a timely and accurate manner, HDM uses the RSS indication. If the RSS decays consistently, then the HDM decides to handover to an upper layer network. In order to detect this consistent signal strength decay, the HDM facilitates from a timer mechanism similar to the one used in downward vertical handover. If the currently serving network's RSS is less than the predefined threshold for a certain amount of time greater than or equal to a threshold timer value, then the HDM decides to handover to an upper overlay network. Here, the threshold timer value is again determined by the application running on MN. That is, the threshold timer value depends on the type of application traffic. We can simply classify the traffic as real-time and non-real time. In the case of delay sensitive real-time services, handover should be performed as rapidly as possible in order to minimize the delay due to frequent handovers. For non-real time services, the amount of transmission data is more important than the delay. Therefore, the connection to the underlay should be maintained as long as possible. Thus, the threshold timer value for real-time services should be much more smaller than that of non-real time services.

Vertical Handover Execution Phase. This section proposes different signaling flows for downward and upward vertical handover. These signaling flows are inspired by the flows demonstrated in [42].

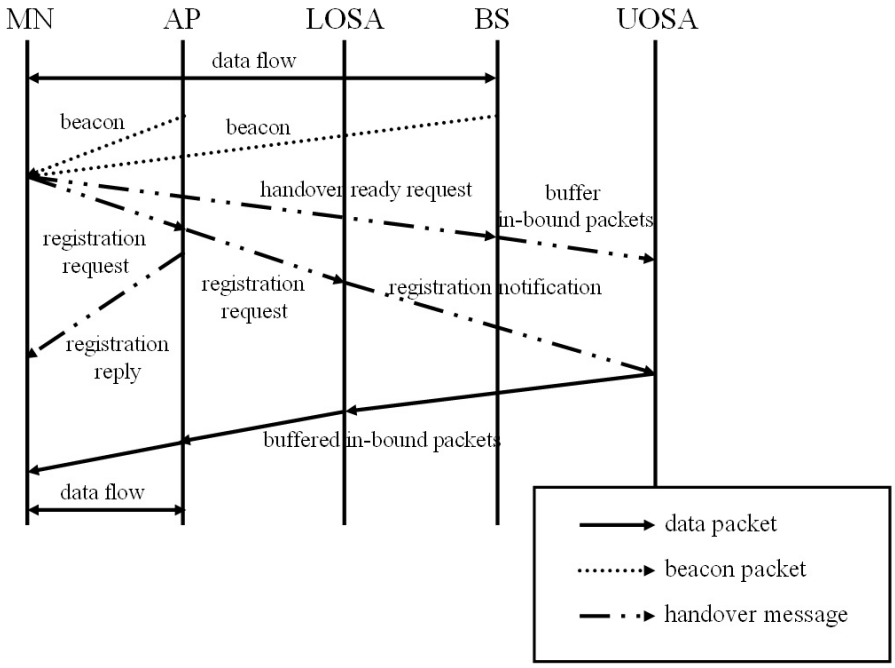


Figure 5.6. Downward vertical handover signaling flow

When an MN enters and connects an underlay wireless network, it performs a downward vertical handover. Fig. 5.6 depicts the proposed signaling flow for downward vertical handover. For the sake of simplicity, downward vertical handover between a WWAN and WLAN is considered in this figure without loss of generality. In this case, power saving can be achieved by determining the time of the physical layer sensing in the handover transition region, and the time to activate the underlay card in the MN. For instance, position information can be used. For the seamless handover service, the handover point in the downward vertical handover flow is not a critical factor, because the upper overlay covers the underlay region.

In Fig. 5.6, the HDM residing on the MN performs physical layer sensing to detect availability of any WLAN. If there is an available WLAN, the WLAN card is activated. After that, the utility of WLAN is checked by HDM in order to determine whether to make a handover or not. If the conditions for the handover are satisfied, MN is informed by HDM and the handover procedure starts.

First, the MN sends a *Handover Ready Request* to the currently serving BS. The BS warns its UOSA to buffer all in-bound packets coming to the MN during the

handover. Thus, all the in-bound packets coming to the MN will be directly sent to the UOSA to be buffered. Then, the MN sends a *Registration Request* to the AP in the WLAN. If the AP accepts this registration request, it sends a positive reply back to MN and forwards this request with an *accepted sign* to its LOSA. The LOSA sends the related notification to the UOSA. By this way, after registration, the UOSA can send all the buffered in-bound packets to the MN via LOSA and AP. Finally, the MN starts communicating with the WLAN. If the AP does not accept the MN's registration request, it sends a negative reply back to MN and forwards this request with a *rejected sign* to its LOSA. Thus, the LOSA sends the related notification to the UOSA and the UOSA stops buffering in-bound packets.

When the MN enters and connects to an upper overlay region, it performs an upward vertical handover. Fig. 5.7 depicts the proposed signaling flow for upward vertical handover. For the sake of simplicity, upward vertical handover between a WLAN and WWAN is considered in this figure without loss of generality. As the MN leaves the WLAN AP, the strength of the beacon signal that is received from the AP periodically weakens. If its strength decreases below the threshold value, then the WWAN card is activated and starts to make signaling with the system to prepare the handover. When the HDM receives continuous beacon signals below the threshold value, which is sensed by timer mechanism of HDM, the handover execution process is performed.

When the MN is warned by HDM, it sends a *Handover Ready Request* to the currently serving AP. The AP warns its LOSA to buffer all in-bound packets coming to the MN during the handover. Thus, all the in-bound packets coming to the MN will be directly sent to the LOSA to be buffered. Then, the MN sends a *Registration Request* to the BS in the WWAN. If the BS accepts this registration request, it sends a positive reply back to MN and forwards this registration request to its UOSA. The UOSA sends the related notification to the LOSA. By this way, after registration, the LOSA can send all the buffered in-bound packets to the MN via UOSA and BS. Finally, the MN starts communicating with the WWAN. If the BS does not accept the MN's registration request, it sends a negative reply back to MN and forwards this request

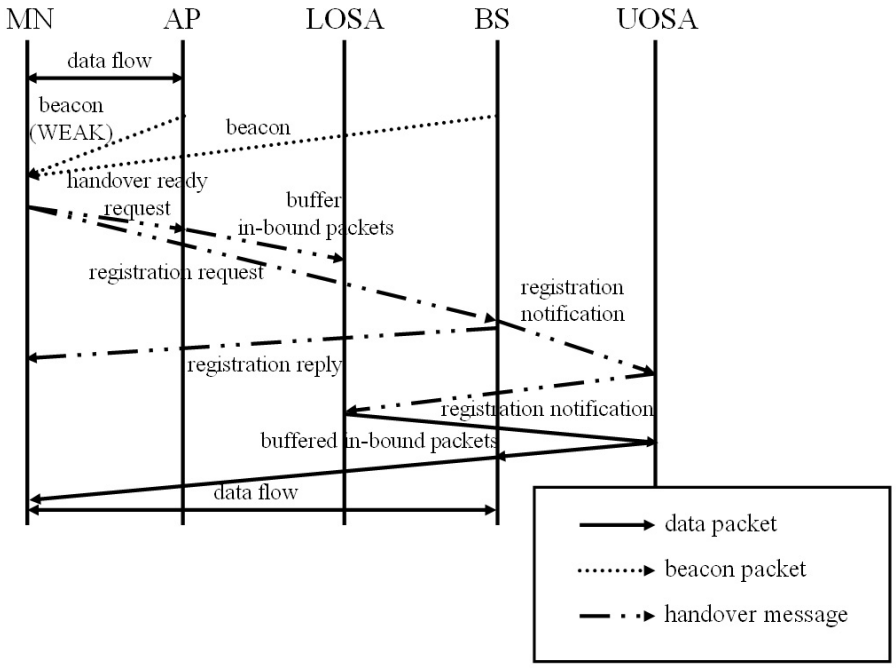


Figure 5.7. Upward vertical handover signaling flow

with a *rejected sign* to its UOSA. Thus, the UOSA sends the related notification to the LOSA and the LOSA stops buffering in-bound packets. Table 5.2 lists the messages used in the signaling flows.

Table 5.2. Message lengths for vertical handover

Message	Length (bits)	Content
Handover Ready Req.	128	MN's IP
Buffer Inbound Packets	129	MN's IP + buffer sign
Reg. Req.	128	MN's IP
Reg. Reply	128	AP's IP
Reg. Not. (1)	256	MN's and AP's IP
Reg. Not. (2)	257	MN's and AP's IP + buffer sign

Improvements in Connection Quality and Billboard Manager. Since a MN notifies its peer directly of its IP address changes, there may be problems in making applications transparent to mobility. Here, transparency to mobility means an upper layer application's being unaware of changes in address, port, and routing-related information. Transparency is very important, especially for those UDP applications

that use the IP address and port number of the received packet in identifying to which user a packet belongs.

The following example illustrates such a transparency problem. For example, suppose that node A with IP address IP_A is communicating with its peer B. The node B buffers the address and port of A and uses it as the identity of A. When A moves to a new place and gets a new IP address, IP_A', the following two things happen:

- When A sends a packet to B using the new IP address IP_A', B cannot know that it is a packet from A; thus, the packet cannot be processed correctly.
- When B sends a packet to A, it still uses the old IP address IP_A. Therefore, the packet will be delivered to the wrong receiver.

Therefore, a Billboard Manager (BM), which resides in the SAs, is proposed to maintain connection's continuity using an end-to-end argument when handover occurs. By utilizing the information provided by the HDM, the BM not only maintains the connection, but also achieves the best possible communication quality. Simply, the BM maintains a map between the original connection information and the current connection information for each active connection. Therefore, upper layer applications become transparent to address changes.

The original connection information does not change during the connection's lifetime, while the current connection information changes each time the MN or its peer gets a new IP address. The upper layer applications only see the original connection information. Besides, a new entry is inserted into the billboard when a new connection is established, and the entry is linked with the current connection information when the source or destination nodes move to a new wireless network interface. Therefore, mobility is made transparent to the upper-layer applications.

5.2.3. Performance Evaluation

In order to evaluate the performance of the proposed vertical handover scheme, both the HDM Algorithm and the SA-based architecture are compared to some prior work. The HDM algorithm is compared to the conventional two-level thresholding scheme which is used in OmniCon [31] and the utility-based scheme similar to the one proposed in [12]. Each algorithm is evaluated by analysing the average number of handovers for different MN speeds, the algorithm's cost of calculation and the MNs' average throughput. Then, the proposed SA-based architecture is compared to the hierarchical architecture in [25] and the architecture proposed in [27]. Each system is evaluated by analysing the breakdown of handover latency and signaling overhead.

The performance of the vertical handover schemes are evaluated by a C++ simulation. The simulation environment is a 7.5 km × 3 km area and consists of 61 WLAN access points (AP) with 300 m of coverage and 2 GSM base stations (BS) with 5 km of coverage. Thus, the WON implemented in this simulation consists of 2 layers. Fig. 5.8 depicts the map of the simulation environment. The dots represent the WLAN APs whereas the triangles represent the BSs.

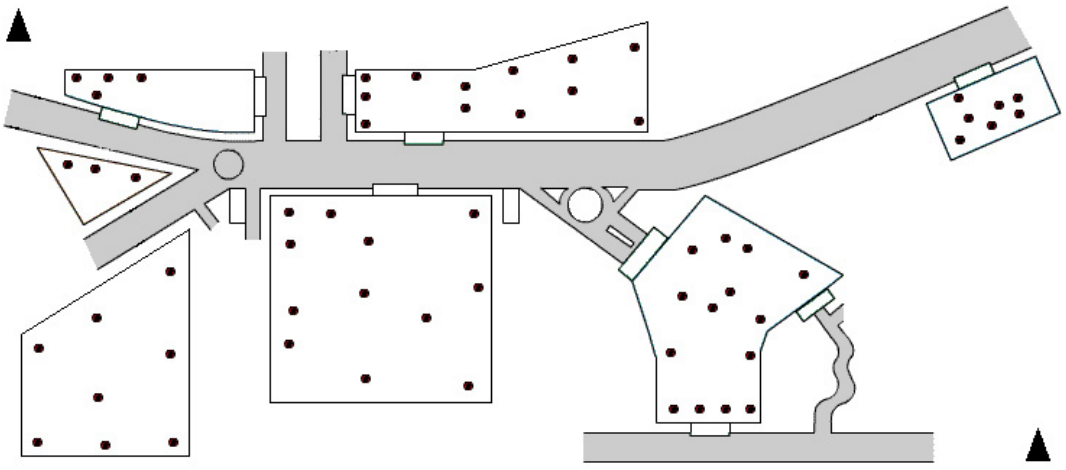


Figure 5.8. Map of the simulation environment

The channel propagation model is taken from [33]. In this model, RSS at distance d is calculated according to Equation 5.2:

$$RSS(d) = P_t - PL(d) + X_\sigma \tag{5.2}$$

where P_t is the transmitted power, PL(d) is the path loss at distance d and X_σ is a zero-mean Gaussian random variable with standard deviation σ. Here, PL(d) is given by Equation 5.3:

$$PL(d) = S + 10 \cdot n \cdot \log(d) \tag{5.3}$$

where S is the path loss constant depending on the propagation environment and n is the path loss exponent.

Table 5.3 depicts the simulation parameter values for access providers. The values for n, σ and S are taken from [33] whereas the other values are taken from commercial products' datasheets such as [43] and [44].

Table 5.3. Simulation parameter values for access providers

Parameter	AP	BS
coverage (m)	300	5000
normalized cost	0.1	1
bit rate (kbps)	11000	100
P_t (mW)	63	25000
n	3.3	4
σ (dB)	7	6
S (dBm)	28.7	19

The environment contains 500 MNs moving randomly for 3 hours within the map borders. The MNs choose one of the four directions: northwest, northeast, southwest and southeast. They move a random amount of distance horizontally and vertically in the chosen direction. An MN changes its direction whenever it hits one of the borders. Such a trajectory generation asserts the MNs to move larger distances and come across a greater number of candidate wireless networks.

Each MN can either do web browsing or video conferencing. The type of application running on MN affects the parameters used in the HDM algorithm described earlier. Table 5.4 lists the values for these parameters.

Table 5.4. Simulation parameter values for HDM algorithm

Parameter	Web Browsing	Video Conferencing	Section
w_B	2	10	I
w_{RSS}	5	5	
w_C	10	3	
w_L	3	8	
$T_{timerDOWN}$ (s)	8	8	
T_{RSSLOW} (mW)	RSS(0.8 × coverage)	RSS(0.8 × coverage)	II
$T_{RSSHIGH}$ (mW)	RSS(0.5 × coverage)	RSS(0.5 × coverage)	
$T_{timerUP}$ (s)	5	5	III
T_{RSSUP} (mW)	RSS(0.8 × coverage)	RSS(0.8 × coverage)	

Parameters in Section I of Table 5.4 are used in downward vertical handover decisions. w_B, w_{RSS}, w_C and w_L are the weights of network's bitrate, RSS, network's cost and load, respectively. $T_{timerDOWN}$ is the threshold timer value for monitoring the consistent utility improvement in the target wireless network. That is, if utility of the target network is greater than the utility of the current network for at least 8 s, then the HDM decides to perform a downward vertical handover to improve QoS.

Parameters in Section II of Table 5.4 are used in horizontal handover decisions. Parameters T_{RSSLOW} and $T_{RSSHIGH}$ are the lower and higher thresholds of the two-level thresholding scheme, respectively. T_{RSSLOW} corresponds to the RSS at distance (0.8 × coverage) whereas $T_{RSSHIGH}$ corresponds to the RSS at distance (0.5 × coverage).

Finally, parameters in Section III of Table 5.4 are used in upward vertical handover decisions. $T_{timerUP}$ is the threshold timer value for monitoring the consistent RSS decay. That is, if RSS is less than T_{RSSUP} for 5 s, then the HDM decides to make an upward vertical handover. Here, T_{RSSUP} corresponds to the RSS at distance (0.8 × coverage). With reference to the video traces analysed in [45], the average video frame length is taken as 5 s. Thus, any disruption continuing for 5 s will degrade QoS and help

us to determine a consistent RSS decay. The results of an Internet traffic modelling may be considered to find an accurate $T_{timerUP}$ value for web browsing. Since this is out of the scope of this thesis, the same $T_{timerUP}$ value is assumed for web browsing.

Performance Evaluation of HDM. Fig. 5.9 depicts the average number of handovers an MN performs as it moves faster. The MN's velocity changes between 1 m/s and 20 m/s. Among the three handover decision algorithms, two-level thresholding scheme has the poorest performance since it leads to a large number of handovers. The algorithm only considers the relative strengths of received signals and decides a handover whenever the threshold values are met. Since it does not consider the MN's speed, it may mistakenly allow fast moving MNs to be served by lower overlays. A careful selection of threshold values and addition of a timer mechanism may prevent possible oscillations between wireless networks of different overlays.

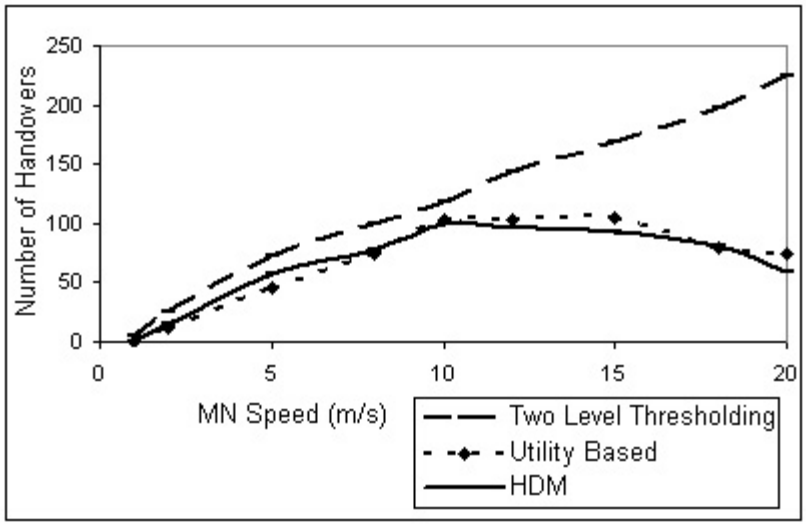


Figure 5.9. Average number of handovers an MN performs v.s. MN's speed

On the other hand, the utility-based and HDM algorithms result in a much lower number of handovers. The performance of these two algorithms are almost the same since they calculate the same utility function during downward vertical handovers. Therefore, we can conclude that applying two-level thresholding in horizontal handovers and monitoring RSS decay in upward vertical handovers are as reliable as using utility functions.

Fig. 5.10 compares the calculation cost of each decision algorithm as the MNs move with 10 m/s. Note that, the costs are considered only when the MNs' speed is 10 m/s because the calculation costs do not change drastically with the MN's mobility. In order to find these costs, certain cost values are assigned to the operations shown in Table 5.5. The operation costs are determined according to the instruction execution times of various operations listed in [46].

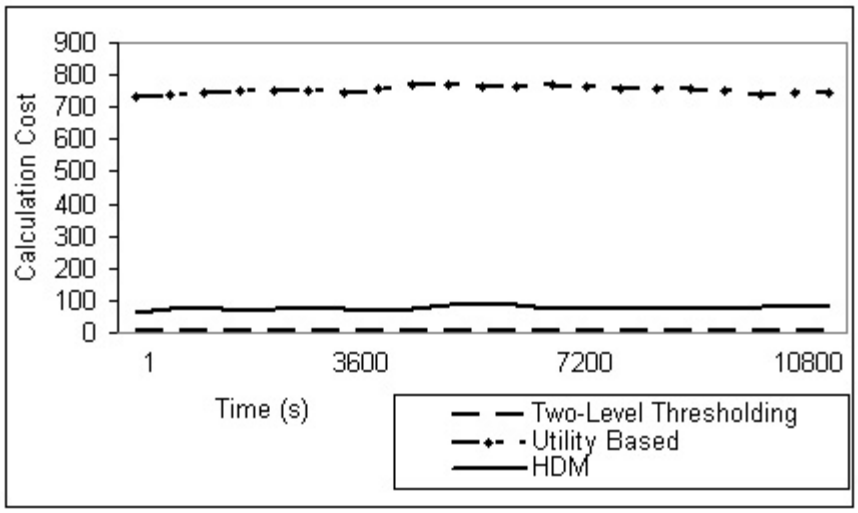


Figure 5.10. Cost of handover decision calculation

As depicted in Fig. 5.10, two-level thresholding has the lowest calculation cost since it is very simple and does a mere comparison operation. On the other hand, utility-based algorithm has a high cost since it does a series of multiplications and additions to consider all the factors affecting the utility of the network (e.g. MN's speed, network's coverage, cost, bitrate, etc.). Finally, the HDM algorithm appears to have an intermediate cost since it performs different tasks for different types of handovers. Calculating utility is usually an overkill for deciding horizontal and upward vertical handovers.

Finally, Fig. 5.11 compares the MNs' average throughput as they move faster. Since the two-level thresholding scheme lets an MN to handover immediately, the resulting throughput is relatively higher. The other two schemes do not let an immense number of handovers and do not let fast moving MNs to communicate with lower overlays unneccessarily. Therefore, they results in a bit lower but more steady throughput.

Table 5.5. Cost of operations

Operation	Cost
addition	3
multiplication	35
division	70
comparison	3

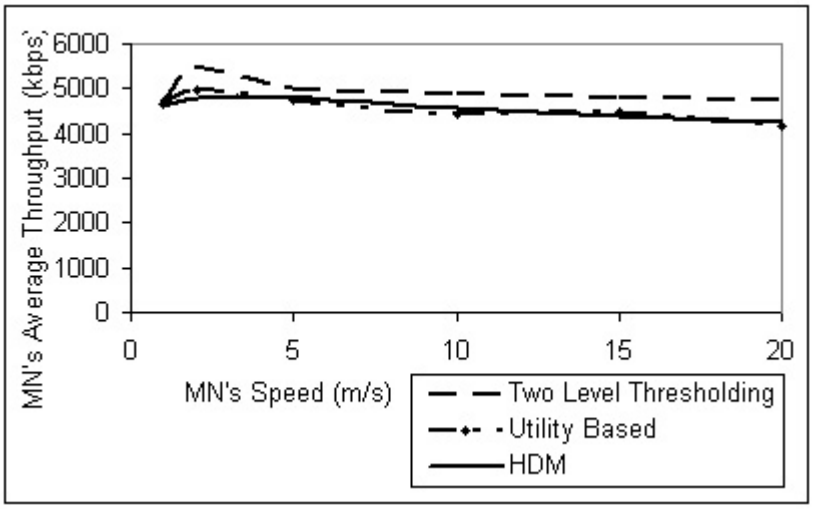


Figure 5.11. MNs' average throughput v.s. MNs' speed

Performance Evaluation of SA-Based Architecture. This section analyses the breakdown of handover latency and signaling overhead. By this way, it compares the proposed SA-based architecture against the architectures in [25] and [27].

Badis et al. proposes a hierarchical architecture for reducing handover latency [25]. The proposed architecture is shown in Fig. 5.12. The Gateway Foreign Agent router implements the role of a FA as well as the role of a HA if the MN is in its home network. It periodically broadcasts agent advertisement messages containing its IP address. The MN registers its Gateway Foreign Agent router care-of-address when it enters a foreign subnet. By this way, when the MN is away from its home network, the HA intercepts packets addressed to it and sends them to the Gateway Foreign Agent router. The Gateway Foreign Agent router decapsulates packets sent by the HA and forwards them to the MN.

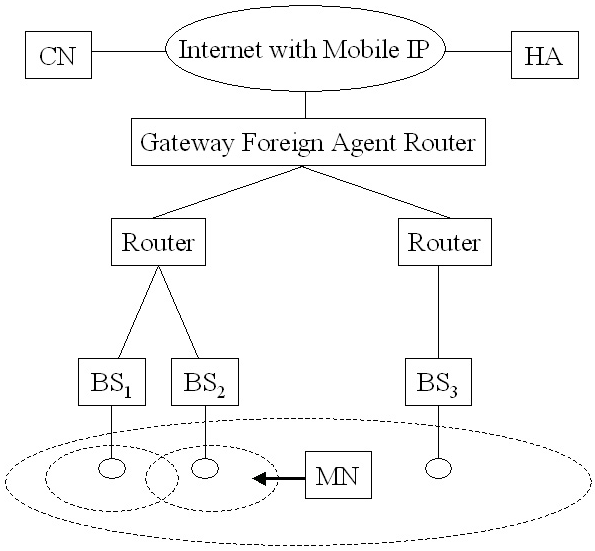


Figure 5.12. Badis' hierarchical mobility management

BSs send periodic agent advertisement messages. By listening to these signals, the MN determines which BS should be the forwarding one. The MN keeps track of the IDs of the last few packets it has received. During a handoff, the MN sends the new BS a *Registration Request* message together with this list of IDs. The new BS sends the old BS a *Notification* message together with this list. By this way, the old BS resends only those buffered packets that are not in the list of IDs. The architecture assumes that all the signaling between the BSs are carried out by a wired backbone. Badis' signaling flow is not clearly presented in [25]. However, by carefully reading [25], one can draw a simple signaling flow by himself as depicted in Fig. 5.13.

As depicted in Fig. 5.13, the handover is made between two base stations of the same layer. Hence, it is a horizontal handover. Throughout the paper, Badis et al. describes horizontal handover by making use of the upper overlays within a WON. Moreover, it does not provide detailed information about the structure and length of the messages. Therefore, [25] cannot be a proper reference point for comparing signaling overhead and latency of a vertical handover.

Nevertheless, one can calculate the signaling overhead by the approximate message lengths listed in Table 5.6 in order to get a slight comprehension of this architecture. Such a calculation will yield an overhead of (128 + 128) + (256 + 128) + 2 x

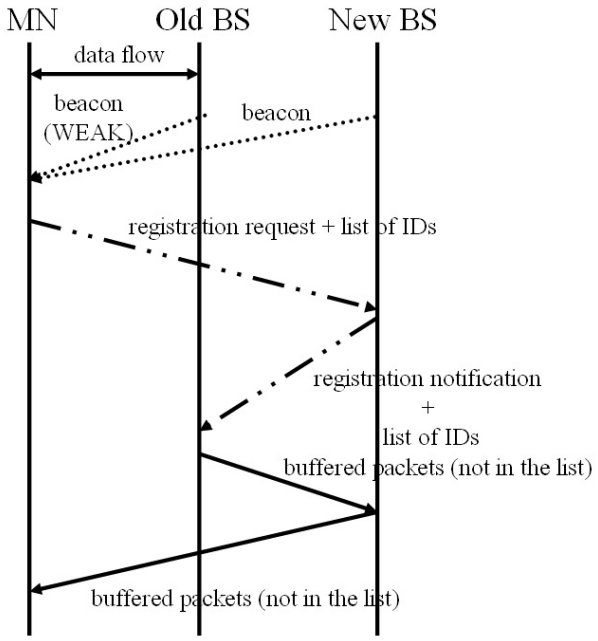


Figure 5.13. Badis' handover signaling flow

n_{buf} = 640 + 2 x n_{buf} bits where n_{buf} is the number of buffered packets. Here, n_{buf} is an unknown but definitely proportional with the handover latency.

Table 5.6. Length of messages in Badis's architecture

Message	Length (bits)	Content
registration request	128	MN's IP
registration notification	256	MN's and new BS's IP
list of IDs	128	8 x 16-bit IDs

Wang et al. proposes an integrated architecture for UMTS and IEEE802.11 WLAN [27] as depicted in Fig. 5.14. The proposed handover signaling flow changes according to the networks' administrative domains (or operators). When the networks are in the same administrative domain, the signaling flow for handover is simpler and hence has a smaller latency and overhead. However, when the networks are in different domains, the latency and overhead increases due to the extra messaging between the architecture's components. Fig. 5.15 depicts these different signaling flows.

According to the message lengths defined in [27] and listed in Table 5.7, the total overhead is calculated to be 128 + 129 + 2 x 257 + 2 x 257 = 1285 bits when the

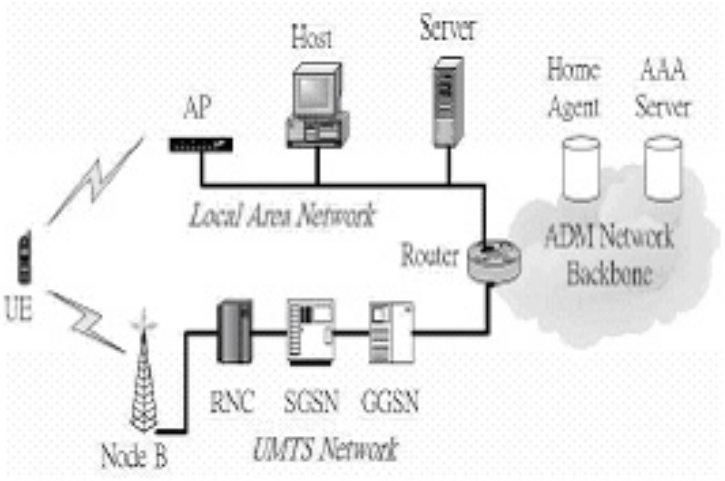


Figure 5.14. Wang's integrated UMTS-WLAN architecture

networks are in the same domain. When they are in different domains, this overhead is calculated to be 128 + 129 + 128 + 128 + 2 x 128 + 2 x 128 + 2 x 257 + 2 x 257 = 2053 bits.

Table 5.7. Length of messages in Wang's architecture

Message	Length (bits)
vertical handover request	128
vertical handover reply	129
routing table update message	257
routing table update acknowledgement	257
router solicitation	128
router advertisement	128
binding update	128
binding acknowledgement	128

As for the architecture proposed in this thesis, the total overhead due to horizontal handover messaging is 256 + 128 + 128 + 128 + 128 = 768 bits. For vertical handover, the messaging overhead is 128 + 129 + 128 + 128 + 256 + 257 = 1026 bits regardless of the administrative domains. Note that, the message lengths used in this calculation are derived from Tables 5.1 and 5.2. Therefore, the proposed SA based architecture has a lower horizontal and vertical handover overhead than that of Badis' architecture for $n_{buf} \geq 2$ assuming that each buffered packet is 128-bits. Note that both the horizontal

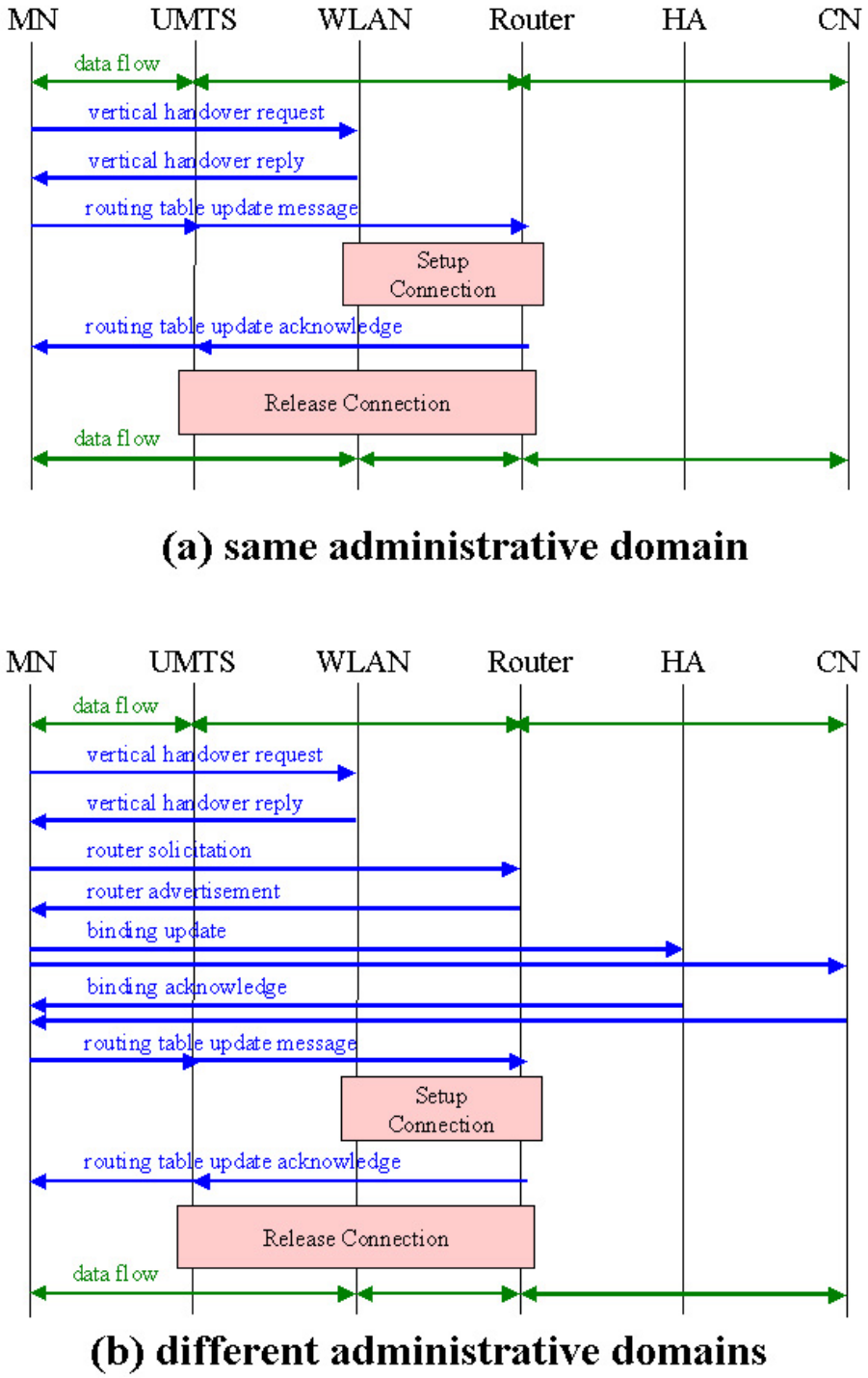


Figure 5.15. Handover from UMTS to WLAN in Wang's architecture

and vertical handover overheads are compared against Badis's architecture overhead since Badis et al. claims to perform vertical handovers but presents only horizontal handovers. Finally, the proposed SA-based architecture definitely has a lower vertical handover overhead than that of Wang's architecture regardless of the administrative domains of the networks.

In order to compare handover latencies, the latency of sending a message from one device to another, L, is calculated according to Equation 5.4. This equation is adopted from the analysis made in [25].

$$L = S_{MSG}/BW + L_{NI} \tag{5.4}$$

where S_{MSG} is size of the message in bits, BW is the bandwidth of the network in bits/s and L_{NI} is the latency of the network interface in s.

Without loss of generality, assuming that a MN performs an upward vertical handover, the latencies of each message in each architecture would be as listed in Table 5.8. In this table, L_U, L_L and L_W are the latencies encountered in sending a 128-bit message in the upper wireless network, lower wireless network and wired network, respectively. Note that, in general, $L_U \gg L_L \gg L_W$ since $BW_U \ll BW_L \ll BW_W$ as depicted in Table 2.1.

According to the latencies listed in Table 5.8, upward vertical handover latency in Badis' architecture, L_{Badis}, is approximately

$$L_{Badis} = L_U + 2 \times L_W + n_{buf} \times L_W + n_{buf} \times L_U \tag{5.5}$$

$$L_{Badis} = (1 + n_{buf})L_U + (2 + n_{buf})L_W \tag{5.6}$$

where n_{buf} is the number of packets not in the list of IDs.

In Wang's architecture, if the wireless networks are in the same administrative

Table 5.8. Approximate latency of messages in each architecture

Message	Latency	Section
registration request	L_U	I
registration notification	$2 \times L_W$	
list of IDs (from old BS to new BS)	L_W	
list of IDs (from new BS to MN)	L_U	
vertical handover request	L_U	II
vertical handover reply	L_U	
routing table update message	$L_L + L_W$	
routing table update acknowledgement	$L_L + L_W$	
router solicitation	$L_U + L_W$	
router advertisement	$L_U + L_W$	
binding update	$L_U + 2 \times L_W$	
binding acknowledgement	$L_U + 2 \times L_W$	
Handover Ready Req.	L_L	III
Buffer Inbound Packets	L_W	
Reg. Req.	L_U	
Reg. Reply	L_U	
Reg. Not. (1)	$2 \times L_W$	
Reg. Not. (2)	$2 \times L_W$	

domain, the upward vertical handover latency, L_{Wang1}, is approximately

$$L_{Wang1} = L_U + L_U + (L_L + L_W) + L_{setup} + (L_L + L_W) + L_{release} \tag{5.7}$$

$$L_{Wang1} = 2 \times (L_L + L_U + L_W) + L_{setup} + L_{release} \tag{5.8}$$

where L_{setup} is the connection setup time and $L_{release}$ is the connection release time. If the wireless networks are in different domains, then the upward vertical handover latency, L_{Wang2}, becomes

$$L_{Wang2} = L_U + L_U + (L_U + L_W) + (L_U + L_W)$$

$$+(L_U + 2 \times L_W) + (L_U + 2 \times L_W)$$

$$+(L_L + L_W) + L_{setup} + L_L + L_W + L_{release} \quad (5.9)$$

$$L_{Wang2} = 2 \times L_L + 6 \times L_U + 8 \times L_W + L_{setup} + L_{release} \quad (5.10)$$

Note that [27] does not give any information about the length of L_{setup} or $L_{release}$.

Finally, in the proposed SA-based architecture, the upward vertical handover latency, L_{SA}, is approximately

$$L_{SA} = L_L + L_W + L_U + 2 \times L_W + 2 \times L_W$$

$$+n_{buf} \times L_W + n_{buf} \times L_W + n_{buf} \times L_U \quad (5.11)$$

$$L_{SA} = L_L + (1 + n_{buf})L_U + (5 + 2n_{buf})L_W \quad (5.12)$$

Note that the latency for Registration Reply message is not included in the calculation since it is sent in parallel with Registration Notification (1) message. In the proposed architecture, the majority of the messages are sent via the wired network which has a much greater bandwidth.

In all these latency equations, L_U is the dominant term since the upper layer has the lowest bandwidth. By examining the coefficient of L_U, one can conclude that L_{Badis} and L_{SA} are very close values. Moreover, for $n_{buf} = 1$, L_{Wang1} and L_{SA} become close wheras for $n_{buf} = 5$, L_{Wang2} and L_{SA} become close values. Besides, we have no evidence or idea about neither L_{setup} nor $L_{release}$. Therefore, these latencies depend on actual system parameters and one cannot make a definite conclusion unless experimenting with practical values.

6. CONCLUSION AND FUTURE WORK

In this thesis, handovers in Wireless Overlay Networks (WON) are investigated. A survey of handover schemes lately proposed for WONs is presented and their drawbacks are discussed. Then, by taking into consideration the drawbacks and incompleteness of the current schemes in literature, a new mobility management scheme providing solutions both for horizontal and vertical handovers is proposed. This novel scheme handles the movements of MNs among different networks and divides into two main parts.

The first part covers the problem of horizontal handovers and presents a solution which reduces latency as much as possible. The proposed horizontal handover model decreases latency significantly at the expense of slightly increased traffic. Moreover, by facilitating from a two-level thresholding scheme, it prevents the ping-pong effect. Finally, the simulation results show that the proposed horizontal handover scheme is more appropriate for highly crowded subnets.

The second part covers the problem of vertical handovers. The proposed vertical handover model is based on Handover Decision Manager (HDM) and Subnet Agents (SAs). The proposed HDM scheme gives priority to lower layer candidate networks and performs different tasks for deciding different types of handovers. The algorithm also considers many factors such as the candidate networks cost, coverage, data rate and traffic load. The simulation results show that the proposed HDM decreases the number of unnecessary handovers as well as save power. Besides, the average throughput of the MNs does not decrease significantly as they move faster.

Besides, the SAs manage the handover process to reduce handover latency and to prevent data loss during roaming of the MN between different overlays as well as maintain the connection transparently. In order to provide the connection's continuity in a transparency manner, a Billboard Manager (BM) is suggested. In addition to this, the novel SA-based architecture brings a new signaling flow for downward and upward

vertical handovers. The performance of the architecture is evaluated by comparing its signaling flow with that of the two latest architectures in literature. The analytical work done for this evaluation shows that the proposed SA-based architecture has a significantly lower overhead and an acceptable latency.

The proposed Mobility Management Scheme is a fully-fledged scheme for WONs and still can be improved in many ways. For future work, the simulations can be revised once more. For example, some parameter values in Table 5.4 are determined intuitively. Instead, a sensitivity analysis can be made for parameter value justification. By this way, the simulations may be performed again by the new parameter values. Moreover, more aspects of the architecture should be evaluated such as the probability of blocking, the probability of dropping and the overall complexity. In addition to the improvements in the simulations, some QoS enhancements can be applied to this architecture as described in [36]. Eliminating the burst of losses and hence smoothing losses, packet prioritization, statistical bandwidth and loss guarantees for high-priority applications or users are some of these enhancements.

REFERENCES

1. Andersen, D. G., *Improving End-to-End Availability Using Overlay Networks*, Ph.D., Massachusetts Institute of Technology, February 2005.

2. Brewer, E. and R. Katz, "The Case for Wireless Overlay Networks", *SPIE Multimedia and Networking Conference (MMNC'96), San Jose, CA*, 1996.

3. Vidales, P., G. Mapp, F. Stajano, J. Crowcroft, and C. J. Bernardos, "A Practical Approach for 4G Systems: Deployment of Overlay Networks.", *TRIDENTCOM*, pp. 172–181, 2005.

4. Gurtov, A., *Efficient Data Transport in Wireless Overlay Networks*, Ph.D., University of Helsinki, Finland, April 2004.

5. Porta, T. F. L., K. K. Sabnani, and R. D. Gitlin, "Challenges for nomadic computing: mobility management and wireless communications", *Mob. Netw. Appl.*, Vol. 1, No. 1, pp. 3–16, 1996.

6. Varshney, U. and R. Vetter, "Emerging mobile and wireless networks", *Commun. ACM*, Vol. 43, No. 6, pp. 73–81, 2000.

7. Bernaschi, M., A. Pescapè, F. Cacace, and S. Za, "Analysis and Experimentation over Heterogeneous Wireless Networks.", *TRIDENTCOM*, pp. 182–191, 2005.

8. Stemm, M. and R. H. Katz, "Vertical Handoffs in Wireless Overlay Networks", *Mob. Netw. Appl.*, Vol. 3, No. 4, pp. 335–350, 1999.

9. Badis, H. and K. A. Agha, "An Efficient Mobility Management in Wireless Overlay Networks", *Proceedings of the 14th IEEE 2003 International Symposium on Personal, Indoor and Mobile Radio Communications, San Jose, CA*, Vol. 3, pp. 2500–2504, 2003.

10. Zahariadis, T., "Global Roaming in Next-Generation Networks", *IEEE Commun. Mag.*, Vol. 40, No. 2, pp. 145–151, 2002.

11. Chakravorty, R., P. Vidales, K. Subramanian, I. Pratt, and J. Crowcroft, "Performance Issues with Vertical Handovers - Experiences from GPRS Cellular and WLAN Hot-spots Integration.", *PerCom*, pp. 155–164, 2004.

12. Chen, W.-T., J.-C. Liu, and H.-K. Huang, "An Adaptive Scheme for Vertical Handoff in Wireless Overlay Networks.", *ICPADS*, pp. 541–548, 2004.

13. Johnson, D., C. Perkins, and J. Arkko, "Mobility Support in IPv6", RFC 3775 (Proposed Standard), June 2004, http://www.ietf.org/rfc/rfc3775.txt.

14. Saha, D., A. Mukherjee, I. Misra, and M. Chakraborty, "Mobility Support in IP: A survey of Related Protocols", *IEEE Network*, Vol. 18, No. 6, 2004.

15. Montavont, N. and T. Noël, "Handover Management for Mobile Nodes in IPv6 Networks", *IEEE Commun.*, Vol. 40, No. 8, pp. 38–43, 2002.

16. Liu, S. and S. Yang, "The Impact of DAD on Handoff Performance of Mobile IPv6 and Test of MLD-Based DAD", *Proceedings of the 1st International Conference on Mobile Computing and Ubiquitous Networking, Yokosuka, JAPAN*, pp. 192–195, 2004.

17. Bernaschi, M., F. Cacace, and G. Iannello, "Vertical Handoff Performance in Heterogeneous Networks.", *ICPP Workshops*, pp. 100–107, 2004.

18. Belghoul, F., Y. Moret, and C. Bonnet, "IP-based handover management over heterogeneous wireless networks", *LCN 2003, 28th Annual IEEE Conference on Local Computer Networks,October 20-24, 2003, Bonn, Germany*, October 2003.

19. Seshan, S., *Low-Latency Handoffs in Cellular Data Networks*, Ph.D., University of California, December 1995.

20. Perkins, C., "IP Mobility Support Draft 12", 1995.

21. Varshney, U. and R. Jain, "Issues in Emerging 4G Wireless Networks", *Computer*, Vol. 34, No. 6, pp. 94–96, 2001.

22. Murray, K., R. Mathur, and D. Pesch, "Intelligent access and mobility management in heterogeneous wireless networks using policy", *ISICT '03: Proceedings of the 1st international symposium on Information and communication technologies*, pp. 181–186, Trinity College Dublin, 2003.

23. Murray, K. and D. Pesch, "Intelligent Network Access and Inter-system Handover Control in Heterogeneous Wireless Networks for Smart Space Environments", *1st International Symposium on Wireless Communication Systems*, pp. 66–70, 2004.

24. Seshan, S., H. Balakrishnan, and R. H. Katz, "Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience", *Kluwer International Journal on Wireless Personal Communications*, 1997.

25. Badis, H. and K. A. Agha, "A New Technique for Reducing Vertical Handoff Latency in Wireless Overlay Networks", 2004.

26. Buddhikot, M. M., G. P. Chandranmenon, S.-J. Han, Y.-W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and Third-Generation Wireless Data Networks", *INFOCOM*, 2003.

27. Wang, Y.-H., C.-P. Hsu, and J.-Y. Lai, "A Mobile IPv6 Based Seamless Handoff Strategy for Heterogeneous Wireless Networks.", *CIT*, pp. 600–605, 2004.

28. Pahlavan, K., P. Krishnamurthy, A. Hatami, M. Ylianttila, J. P. Makela, R. Pichna, and J. Vallstron, "Handoff in Hybrid Mobile Data Networks", *IEEE Personal Communications*, April 2000.

29. Zhang, Q., Z. G. C. Guo, and W. Zhu, "Efficient mobility management for vertical handoff between WWAN and WLAN", *IEEE Commun. Mag.*, Vol. 41, No. 11, pp.

102–108, 2003.

30. Inayat, R., R. Aibara, and K. Nishimura, "A Seamless Handoff for Dual-interfaced Mobile Devices in Hybrid Wireless Access Networks.", *AINA (1)*, pp. 373–378, 2004.

31. Sharma, S., I. Baek, Y. Dodia, and T. cker Chiueh, "OmniCon: A Mobile IP-Based Vertical Handoff System for Wireless LAN and GPRS Links.", *ICPP Workshops*, pp. 330–337, 2004.

32. Dang, M. S., A. Prakash, D. K. Anvekar, M. Kapoor, and R. Shorey, "Fuzzy Logic Based Handoff In Wireless Networks", *Proceedings of the IEEE 51st Vehicular Technology Conference*, 3, pp. 2375–2379, May 2000.

33. Majlesi, A. and B. H. Khalaj, "An Adaptive Fuzzy Logic Based Handoff Algorithm for Interworking Between WLANs and Mobile Networks", *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Vol. 5, pp. 2446–2451, September 2002.

34. Ylianttila, M., *Vertical Handoff and Mobility – System Architechture and Transition Analysis*, Ph.D., University of Oulu, May 2005.

35. Katz, R. H., E. A. Brewer, E. Amir, H. Balakrishnan, A. Fox, S. Gribble, T. Hodes, D. Jiang, G. T. Nguyen, V. Padmanabhan, and M. Stemm, "The Bay Area Research Wireless Access Network (BARWAN)", *COMPCON '96: Proceedings of the 41st IEEE International Computer Conference*, p. 15, IEEE Computer Society, Washington, DC, USA, 1996.

36. Subramanian, L., I. Stoica, H. Balakrishnan, and R. Katz, "OverQoS: An Overlay Based Architecture for Enhancing Internet QoS", *1st Symposium on Networked Systems Design and Implementation (NSDI)*, San Francisco, CA, March 2004.

37. Chakravorty, R., P. Vidales, L. Patanapongpibul, K. Subramanian, I. Pratt, and

J. Crowcroft, "On Inter-network Handover Performance using Mobile IPv6", 2003, http://citeseer.ist.psu.edu/chakravorty03internetwork.html.

38. Tüysüz, A. and M. Yıldırım, "A New Algorithm for Horizontal Handover Management in Wireless Mobile Networks", *20th International Symposium on Computer and Information Sciences*, 2005.

39. Tüysüz, A., M. Yıldırım, B. Kurumahmut, and F. Alagöz, "Handovers in Wireless Overlay Networks Based on Mobile IPv6", *Akademik Bilişim*, 2005.

40. Tozer, T. C. and D. Grace, "High-altitude Platforms for Wireless Communications", *Electronics and Communication Engineering Journal*, pp. 127–137, June 2001.

41. Yıldırım, M. and F. Alagöz, "A New Handover Decision Management Algorithm in Wireless Overlay Networks", 2006, submitted to Globecom'06.

42. McNair, J., I. F. Akyildiz, and M. D. Bender, "An Inter-system Handoff Technique for the IMT-2000 System", *The Conference on Computer Communications - IEEE INFOCOM 2000*, 1, pp. 208–216, March 2000.

43. "Linksys WAG54G v2 (Wireless-G ADSL Gateway) Datasheet", 2005, http://www-uk.linksys.com.

44. "Contela BTS Specification", Contela, Inc, 2006, http://www.contela.com/product/btsspec.pdf.

45. Alagöz, F., "Integrating the Scene-Length Characteristics of MPEG Video Bitstreams into the DBS Networks", *International Journal of Satellite Communications and Networking*, Vol. 22, No. 2, pp. 217–229, March/April 2004.

46. "MC68000 16-Bit MP User's Manual", Motorola Semiconductor Products Inc., 1980.