M. Tech. (Computer Science) Dissertation Series

Combined Vertical Handoff Decision Algorithm & Call Admission Control in WLAN-UMTS Hybrid Networks

a dissertation submitted in partial fulfillment of the requirement for the M. Tech. (Computer Science) degree of the Indian Statistical Institute

By

Pritam Kumar Paul

M.Tech(CS) - 1105

under the supervision of

Dr. Sasthi C. Ghosh

INDIAN STATISTICAL INSTITUTE

203, Barackpore Trunk Road

Calcutta - 700 035

Indian Statistical Institute

203, B.T. Road. Kolkata : 700108

CERTIFICATE

I certify that I have read the thesis titled **“Combined Vertical handoff Decision and Call Admission Control Scheme in WLAN-UMTS Hybrid Networks”**, prepared under my guidance by Pritam Kumar Paul, and in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Master of Technology in Computer Science of the Indian Statistical Institute.

Dr. Sasthi C. Ghosh
Assistant Professor
Advanced Computing and Microelectronics Unit
Indian Statistical Institute

Kolkata
July, 2013.

Declaration

I, Pritam Kumar Paul, declare that this thesis titled, "Combined Vertical Handoff Decision and Call Admission Control Scheme in WLAN-UMTS Hybrid Networks" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for master degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________________

Date: ______________________________________

Abstract

With the advent of new generation wireless networks and data networks with different coverage and bandwidth, the users are being pushed towards universal wireless access and ubiquitous computing through seamless personal and terminal mobility. Also there is a need of a mechanism that will enable the users perform the handover, that means switch from one network to another, with minimum or no disruption to the ongoing session/call that the user may be engaged in. By different data networks, we mean the different networks like UMTS(802.3),WLAN(802.11) & WiMax(802.16). There are quite a few literatures regarding handovers in-between these networks, out of which we will mainly be concentrating in WLAN-UMTS handovers. Traditional criteria of handovers and some in the literatures suffer from the fact that sometimes the handover decision is postponed for too long resulting in call drop or sometimes its taken too early resulting in some unnecessary handoffs. Here we try to take the decision a bit early such that when the user reach the cut-off point, its ready with the best candidate network to go to. Also, before taking the decision for any handover, we try to eliminate some handover cases by denoting them as unnecessary or superfluous, so that we need not unnecessarily do the rest of the process for cases where it can be avoided. In the other part of this work, apart from being able to take decide about the handover candidate network at appropriate time, we also try to maximize the call admission rate in that candidate network. So, not only we try to minimize the overhead of calculating or taking decisions for unnecessary handover cases, but we switch to the network providing us with the best possible call admission possibility. All these will be discussed in the following sections.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Sasthi C. Ghosh, for providing me the courage to go along with this work. It is because of his patience, advice, encouragement and help that I have been able to present this work.

I would also like to thank all the respected professors of ISI, Kolkata for making my educational experience exciting, enjoyable and helping me to gain a better outlook on different topics of Computer Science.

Finally, I want to thank my friends and most importantly my parents. Its because of their encouragement and believe in me that I have been able to come the way all along. So my heartfelt thanks go to them.

Contents

Chapter 1

Introduction

1.1 Background

With the world moving towards ubiquitous computing, the manufacturers of mobile and handheld devices have started making devices with dual or multiple mode of connectivity. By virtue of that the users, nowadays are neither restricted to use their devices, especially mobile phones, for only making calls or send texts nor they are compelled to use any one technology throughout their communication session. They can freely choose, on their own discretion which technology to use for any particular apps that they are running which can be termed as optional cases. And sometimes it might be necessary for them to switch over while going out of a certain coverage area or while getting very poor signal strength, which can be termed as necessary case. Now, as a particular area can be covered under multiple access technologies, hence when performing a handover the user is presented or faced with more than one choice most of the times. So, a strategy is necessary the will facilitate the user in taking the decision keeping in mind its various requirements and above all making sure the ongoing session of the user is less or not disrupted at all. This is where perhaps a handover decision algorithm comes into effect. It makes the process of choosing a network much easier and makes sure the required QoS has been made. But, only choosing the right candidate to switch over to is not the full solution of the problem. The algorithm needs to make sure that, after the switch, the ongoing session/call of the user doesn't get dropped. But its not always possible to guarantee that the handoff calls will succeed at the destination network, hence the usual method used is to switch to that network that will give the maximum acceptance probability depending on some criteria. This is termed as call admission control(CaC).

1.2 Scope

The present work deals with a specific type of Vertical Handover,WLAN to UMTS handover(not the other way round).The work first focuses on eliminating some handovers by terming them as unnecessary or superfluous. The advantage of which is reduction of system overhead by a certain proportion. Apart from that, we try to answer the question that - 'when should we start the handover decision process' to make sure that we are ready to take the decision well before the ultimate cut-off point while not taking the decision too early nor leaving it for too late. Also part of the present work, a combined Call Admission Control scheme, that is overlapped with the later phases of handover decision algorithm. This will enable the mobile user, which is going to be handed off, with the choice of the base station which will provide maximum call admission probability among all the available base stations. All these are explained in the following sections, starting of the relevant issues in Handover and CaC, followed by the tools, techniques and methodologies and our proposed method and finally some results.

1.3 Relevant Works

There has been quite a few study going on regarding this vertical handoff matter. Some of them estimate the problem by modeling them by a cost function and solving like an optimization problem while some use neural network or some other machine learning techniques. In [1] a cost function is developed for multiservice networks that considers several factors in a two-dimensional handoff cost function. In [2] the authors consider handoffs between a WLAN and a cellular network. The handoff policy is different, depending on the nature of the user traffic. For example, since the data rate provided by WLAN (1–10 Mb/s) is much larger than that provided by the cellular network (9.2–200kb/s), the handoff policy for non-real-time services is to attempt to use the services of the WLAN as long as possible. Thus, the preferred handoff point from the cellular network to the WLAN is the first time the signal strength in the WLAN reaches an acceptable level, while the handoff point from the WLAN to the cellular network is the last time the signal strength falls below the acceptable level. In [3] multi-network architectural issues were explored, and an advanced neural-network based vertical handoff algorithm was developed to satisfy user bandwidth requirements. In [4] the network selected is based on Analytic Hierarchy Process (AHP) and Grey Relational Analysis (GRA). AHP decomposes the network selection problem into several subproblems and assigns a weight value for each sub-problem. GRA is then used to rank the candidate networks and selects the one with the highest ranking. In [5] the vertical handoff decision is formulated as a fuzzy MADM (Multiple Attribute Decision Making) problem.

Fuzzy logic is used to represent the imprecise information of some attributes and user preferences. Thus we summarize some recent works on vertical handoff field.

Chapter 2

Issues in Vertical handover and Call Admission Control

2.1 Beyond 3G and moving towards ubiquitous computing

Ubiquitous computing (ubicomp) is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activities. In the course of ordinary activities, someone "using" ubiquitous computing engages some computational devices and systems simultaneously, and may not necessarily even be aware that they are doing so. This model is considered an advancement from the older desktop paradigm. More formally, ubiquitous computing is defined as "machines that fit the human environment instead of forcing humans to enter theirs". This paradigm is also described as pervasive computing, ambient intelligence, or, more recently, everyware, where each term emphasizes slightly different aspects. When primarily concerning the objects involved, it is also physical computing, the Internet of Things, haptic computing, and things that think. Rather than propose a single definition for ubiquitous computing and for these related terms, a taxonomy of properties for ubiquitous computing has been proposed, from which different kinds or flavors of ubiquitous systems and applications can be described.

In mobile and wireless communications, the research community has perceived the limitations of 3G cellular communication systems in terms of user throughput and cost of operation and usage and consequently has focused on research towards systems Beyond Third Generation (B3G) that outside Europe have been called Forth Generation (4G) systems. The recommendation M.1645 of International Telecommunications Union - Radio (ITU- R) has been approved to be the basis of future activities for systems beyond 3G. The ITU vision aims at integration and cooperation of existing and evolving access networks on the one hand and advanced air interfaces with significantly improved performance compared to 3G systems on the other hand. Driven by the EU research framework program, European industry has federated in the Wireless World Initiative (WWI) towards a common technological, industrial, regulatory and service approach to realize this vision.

Realization of this vision demands a major shift from the current concept of "anywhere, anytime" to a new paradigm of "any network, any device, with relevant content and context in a secure and trustworthy manner".

As the next step, 3G Long Term Evolution (LTE) is currently standardized. Based on the different radio interface transmission technology Orthogonal Frequency-Division Multiplexing (OFDM), 3G LTE is aiming to achieve a downlink data rate of three to five times higher compared to HSPA in the same bandwidth and significantly lower latency. For most situations, the cost to evolve a UMTS network to a next generation radio interface will be low compared to the cost of deploying a new network. The reason is that most of the existing infrastructure will remain the same, requiring only major upgrades at a base station and on terminals and inter-working with existing systems is easily achievable as the same core network can be used.

- Ubiquitous network infrastructures and architectures : Next generation infrastructures need to support convergence and interoperability of heterogeneous mobile and broadband network technologies. The main enablers are expected to be flexible and spectrum-efficient mobile broadband access technologies and optimized protocols and routing concepts. Moreover, increased scalability, context-awareness and efficient traffic management is needed to support an order of magnitude increase in the number of connected devices and enabling the emergence of applications that are machine-to-machine or sensor-based and are capable of functioning within a multiplicity of public or private operating environments.

- Optimized control, management and flexibility of the future network infrastructure: New control and management schemes are identified as key research topics towards flexible and cognitive networks. They are expected to enable seamless end-to-end network and service composition and operation across multiple operators and business domains

- Technologies and systems architectures for the future Internet: Activities in this area are aimed at overcoming the expected long term limitations of current Internet capabilities, architecture and protocols. They are driven by the need for generalized mobility, scalability from the perspective of devices, service attributes and application environments and security and trusted domains

2.2 Increased Bandwidth Demand - Need for UMTS-WLAN Internetworking

With the advent of newer technologies in mobile telephony, people nowadays are not restricted to use their mobile phone only for making calls or sending texts. The multipurpose use of a mobile device now include online gaming,video streaming,net browsing,chatting etc. Because of these many uses of a mobile phone the need for very high speed connection is increasing day by day and as a result of that, a very large amount of data traffic is incurred. Apart from that,the applications running on the mobile devices are not just standalone locally running applications any more. More often than not,they need to access the net. And new applications of these kind are coming to the market very often. Hence, because of these bandwidth consuming applications,the bandwidth need is much more compared to that of a few days back.
So, keeping track with the need of the users, the mobile service providers have evolved as well. Starting from 2G which first implemented data services apart from greater penetration levels than their predecessor,to 2.5G which introduced GPRS which is believed to be the starting point of 3G, to 2.75G or EDGE with peak data rate of 238kbps. But still they were insufficient to the needs of the user and hence came 3G with much bigger B/W and much higher data rate. But keeping the cost factor in mind,it was seen to be difficult to meet this traffic demand by wide cellular network like 3G UMTS. Hence using WLAN on dual mode devices(having 2 network interface cards) can be useful. The advantage of WLAN over UMTS is that it provides high data rate in very low cost as compared to UMTS. But with a single drawback being very low coverage area.
Hence, as the user will be moving along different networks, internetworking between them is needed as the main objective of the mobile user need to be fulfilled - which is - *increased data rate at low cost over wider coverage.*
One of the main issues in this aspect is Handover which is discussed next.

2.3 Handover

2.3.1 Definition

In cellular telecommunications, the term handover or handoff refers to the process of transferring an ongoing call or data session from one channel connected to the core network to another. In satellite communications it is the process of transferring satellite control responsibility from one earth station to another without loss or interruption of service.

American english use the term handoff, and this is most commonly used within some American organizations such as 3GPP2 and in American originated technologies such as CDMA2000. In British English the term handover is more common, and is used within international and European organisations such as ITU-T, IETF, ETSI and 3GPP, and standardized within European originated standards such as GSM and UMTS. The term handover is more common than handoff in academic research publications and literature, while handoff is slightly more common within the IEEE and ANSI organizations.

2.3.2 Reasons

In telecommunications there may be different reasons why a handover might be conducted:

- When the phone is moving away from the area covered by one cell and entering the area covered by another cell the call is transferred to the second cell in order to avoid call termination when the phone gets outside the range of the first cell.
- When the capacity for connecting new calls of a given cell is used up and an existing or new call from a phone, which is located in an area overlapped by another cell, is transferred to that cell in order to free-up some capacity in the first cell for other users, who can only be connected to that cell.
- In non-CDMA networks when the channel used by the phone becomes interfered by another phone using the same channel in a different cell, the call is transferred to a different channel in the same cell or to a different channel in another cell in order to avoid the interference.
- Again in non-CDMA networks when the user behavior changes, e.g. when a fast-traveling user, connected to a large, umbrella-type of cell, stops then the call may be transferred to a smaller macro cell or even to a micro cell in order to free capacity on the umbrella cell for other fast-traveling users and to reduce the potential interference to other cells or users (this works in reverse too, when a user is detected to be moving faster than a certain threshold, the call can be transferred to a larger umbrella-type of cell in order to minimize the frequency of the handovers due to this movement).
- In CDMA networks a handover may be induced in order to reduce the interference to a smaller neighboring cell due to the "near-far" effect even when the phone still has an excellent connection to its current cell.

2.3.3 Classification

The most basic form of handover is when a phone call in progress is redirected from its current cell (called source) to a new cell (called target). In terrestrial networks the source and the target cells may be served from two different cell sites or from one and the same cell site (in the latter case the two cells are usually referred to as two sectors on that cell site). Such a handover, in which the source and the target are different cells (even if they are on the same cell site) is called inter-cell handover. The purpose of inter-cell handover is to maintain the call as the subscriber is moving out of the area covered by the source cell and entering the area of the target cell.

A special case is possible, in which the source and the target are one and the same cell and only the used channel is changed during the handover. Such a handover, in which the cell is not changed, is called intra-cell handover. The purpose of intra-cell handover is to change one channel, which may be interfered or fading with a new clearer or less fading channel.

2.3.3.1 Hard Handover - Soft Handover

- Hard Handover : a mobile cellular communication network, a Hard handoff (or Hard handover) is a typical Handoff mechanism in a communication network which is designed to work by first breaking off from the initial connection with a base station before switching to another base station. This is done in order to retain communications in a session for mobile users after incurring a non perceptible and insignificant brief interruption. A Hard handoff is also referred to as “Break-before-Make” handover.

  A brief break off before establishing connection with another base station is referred to as an ”event interruption” and is characteristically too short to be even noticed by users communicating on their mobile devices. Although it is noticeable as a short beep in analog communication systems, it is practicably not identifiable by a mobile device user in setup of digital communication. Also, at any moment in time, a Hard handoff is typically implemented on a single channel of communication. A Hard handoff is relatively cheaper and easier to implement in comparison to another type of Handoff called Soft handoff or “Make-before-Break” handover.

- Soft Handover : Soft handover or soft handoff refers to a feature used by the CDMA and W-CDMA standards, where a cell phone is simultaneously connected to two or more cells (or cell sectors) during a call. If the sectors are from the same physical cell site (a sectorized site), it is referred to as softer handoff. This technique is a form of mobile-assisted handover, for IS-95/CDMA2000 CDMA cell phones continuously make power measurements of a list of neighboring cell sites, and determine whether or not to request or end soft handover with the cell sectors on the list.

Due to the properties of the CDMA signaling scheme, it is possible for a CDMA phone to simultaneously receive signals from two or more radio base stations that are transmitting the same bit stream (using different transmission codes) on the different physical channels in the same frequency bandwidth. If the signal power from two or more radio base stations is nearly the same, the phone receiver can combine the received signals in such a way that the bit stream is decoded much more reliably than if only one base station were transmitting to the subscriber station. If any one of these signals fades significantly, there will be a relatively high probability of having adequate signal strength from one of the other radio base stations.

On the uplink (phone-to-cell-site), all the cell site sectors that are actively supporting a call in soft handover send the bit stream that they receive back to the Radio Network Controller (RNC), along with information about the quality of the received bits. The RNC examines the quality of all these bit streams and dynamically chooses the bit stream with the highest quality. Again, if the signal degrades rapidly, the chance is still good that a strong signal will be available at one of the other cell sectors that is supporting the call in soft handover.

Apart from these, recent research has been going of focusing on soft handover between not only CDMA or WCDMA, but also between technologies like WLAN-UMTS, WLAN-WiMax, WiMax-UMTS etc which are basically called vertical handovers, which will be explained in a while.

2.3.3.2 Vertical - Horizontal Handover

Vertical handover or vertical handoff refers to a network node changing the type of connectivity it uses to access a supporting infrastructure, usually to support node mobility. For example, a suitably equipped laptop might be able to use both a high speed wireless LAN and a cellular technology for Internet access. Wireless LAN connections generally provide higher speeds, while cellular technologies generally provide more ubiquitous coverage. Thus the laptop user might want to use a wireless LAN connection whenever one is available, and to 'fall over' to a cellular connection when the wireless LAN is unavailable. Vertical handovers refer to the automatic fallover from one technology to another in order to maintain communication. This is different from a 'horizontal handover' between different wireless access points that use the same technology in that a vertical handover involves changing the data link layer technology used to access the network.

Vertical handoffs between WLAN and UMTS (CDMA2000) have attracted a great deal of attention in all the research areas of the 4G wireless network, due to the benefit of utilizing the higher bandwidth and lower cost of WLAN as well as better mobility support and larger coverage of UMTS. Vertical handovers among a range of wired and wireless access technologies including WiMAX can be achieved using Media independent handover which is standardized as IEEE 802.21.

Things involved in Vertical Handover

- Dual mode card : To support vertical handover, a mobile terminal needs to have a dual mode card, for example one that can work under both WLAN and UMTS frequency bands and modulation schemes.

- Interworking architecture : For the vertical handover between UMTS and WLAN, there are two main interworking architecture: tight coupling and loose coupling. The tight coupling scheme, which 3GPP adopted, introduces two more elements: WAG (Wireless Access Gateway) and PDG (Packet Data Gateway). So the data transfers from WLAN AP to a Corresponding Node on the internet must go through the Core Network of UMTS. Loose coupling is more used when the WLAN is not operated by cellular operator but any private user. So the data transmitted through WLAN will not go through Cellular Networks.

- Handover metrics : In traditional handovers, such as a handover between cellular networks, the handover decision is based mainly on RSS (Received Signal Strength) in the border region of two cells, and may also be based on call drop rate, etc. for resource management reasons. In vertical handover, the situation is more complex. Two different kinds of wireless networks normally have incomparable signal strength metrics, for example, WLAN compared to UMTS. In, WLAN and UMTS networks both cover an area at the same time. The handover metrics in this situation should include RSS, user preference, network conditions, application types, cost etc.

- Handover decision algorithm : Based on the handover metrics mentioned above, the decision about how and when to switch the interface to which network will be made. Many papers have given reasonable flow charts based on the better service and lower cost, etc. while some others, using fuzzy logic, neuron network or MADM methods to solve the problem.

- Mobility management : When a mobile station transfers a user's session from one network to another, the IP address will change. In order to allow the Corresponding Node that the MS is communicating with to find it correctly and allow the session to continue, Mobility Management is used.. The Mobility Management problem can be solved in different layers, such as the Application Layer, Transport Layer, IP Layer, etc. The most common method is to use SIP (Session Initiation Protocol) and Mobile IP.
- Handoff procedure : The handover procedure specifies the control signaling used to perform the handover and is invoked by the handover decision algorithm.

Apart from the mentioned above, a handover can also be categorized by the necessity of the handover. The two categorizations are called Unnecessary handovers and superfluous handovers.

2.3.3.3 Unnecessary - Superfluous Handover

- Unnecessary Handover : A handover is termed as unnecessary if the total time of the mobile equipment in the present cell coverage area is altogether less than the expected handover latency. In such cases, the user requirements will be over before the user reach the termination point so taking time in dealing with such cases only increase the system overload. Hence developing a handover decision mechanism necessarily involves taking care of the unnecessary handover cases.
- Superfluous handovers: As we have discussed in the last paragraph, reduction of unnecessary handovers is necessary, otherwise it could cause network waste.One other kind of handover cases that may be avoided in practical cases is termed as superfluous handovers. A handover is considered as superfluous when a mobile terminal back to the previous PoA is needed within certain time duration ("ping-pong" effect), and such handovers should be minimized.

Another valid categorization of handovers, that specially applies to the vertical handover cases is as follows: As we have briefly mentioned in the texts above, the main reasons of intending to perform a handover are:

1. Received signal quality too bad
2. Received power level too low
3. Mobile Station-Base Station distance too high.
4. Better cell(power budget, cost effective,relative received power high) etc.

2.3.3.4 Imperative - Optional Handover

- Mandatory/Imperative : The first three causes are known as mandatory or imperative causes, i.e. if one of these causes occurs, a handover is necessary to maintain the call. This may happen because the MS is leaving the coverage area of the serving cell (intercell handover) or because there is a strong interferer using the same channel in another cell (intracell handover). For example, in a vertical handover scenario of suppose WLAN-UMTS networks, a user currently connected to a WLAN Access Point going out of the WLAN service area. Now if the WLAN area falls under the coverage of several UMTS BS's, then any ongoing call of the user necessarily needed to be handed over to any of the Base stations. These handover cases are termed as mandatory/imperative handovers.

- Optional : The fourth cause is an optional one, i.e. the link quality in the serving cell is sufficiently good, but there are neighbor cells with better received level. Though its not necessary for the link quality of this specific call, there is a benefit for overall network performance to handover the call to the better cell: A call in the better cell causes less interference, especially, if power control is applied. To achieve the same received level in the better cell, a smaller transmit power can be used in this cell. Also sometimes the cost of the usage or the available speed/bandwidth may become the deciding factor for a user in choosing a particular network. For example, in a vertical handover scenario of suppose UMTS-WLAN networks, a user currently connected to a UMTS base station is moving towards an area covered by a WLAN Access point also. Hence, when the user enters the WLAN coverage area, he may choose to handing over to the WIFI AP for better speeds or he may still keep the connection with BS alive and keep using it as if no other network exist. These cases, where the handover is not a necessity but depends on the choice of the user are termed as non-imperative/optional handovers.

2.4 Call Admission Control(CaC)

We all know that cellular wireless network providers provide services to the mobile users in coverage areas which are roughly known as cells. When a mobile user arrives or originates in a cell, a Call Admission Control procedure need to be invoked by the Base Station to determine whether or not to admit the call. The procedure relies on a predefined constraint,also called admission constraint, to determine whether or not to accept the call while maintaining the required QoS or service requirements of the existing/ongoing calls/sessions.[10] If the constraint is satisfied, then the decision is taken to admit the call otherwise the calls gets blocked out of the system. The main objective of any call admission scheme is to make sure that the call drop probability as less as possible and accommodate more calls to utilize the resources efficiently.

In general, CAC can be grouped into two categories : parameter based (proactive) and measurement based (reactive). Parameter based admission control schemes use a priori traffic specification to determine the parameters of deterministic or stochastic models. On the other hand,measurement based admission control offers QoS to users,without requiring priori traffic specifications or online policing. It depends on the measurement of actual traffic load in the network in making admission decisions.[11] In our case, we will be basically considering a parameter based or pro-active CaC scheme that will be explained in the following sections.

2.4.1 Importance of CaC after Handover

While introducing CaC in the previous section, it is mentioned that a CaC scheme is needed for any call arriving or originating in a cell. By *originating* we mean it is a new call.Although it is easy to realize the importance of a CaC scheme for new calls, a CaC scheme to maximize the call admission probability of a handover call is much more important. Because, more often than not, handoff calls consist of uninterrupted call/data session of a user and it is much more convenient and efficient to drop new calls rather than dropping a handoff call. Hence, any handoff decision algorithm, that determines the necessity and timing to perform a handoff should be complemented with suitable call admission algorithm so that after successful handoff, the main objective, that is, seamless connectivity can be satisfied.

2.4.2 CaC Policies

There has been extensive study on homogeneous CaC schemes previously. Recently there is a trend of studying CaC for heterogeneous networks. People have derived several metric and conditions based on which they have defined their CaC policies. We are just listing some of them here. For example, Haun Chen et al[12] proposed the guard channel based CAC for homogeneous network based on birth and death process, which results in the increase of the new call blocking probability. In a heterogeneous network, a balance in the resource utilization between different network users is to be provided. The users should be allocated with the network that has the highest probability of providing the best QoS for a particular service request. In [13] a CaC algorithm is described with service differentiations for voice and data traffic in a topology where isolated hotspots are meshed into a larger cellular network. In [14] a bandwidth adaptation scheme based on per flow degradation was proposed for heterogeneous wireless networks by defining a concept called degrade profile. Here, in order to admit calls, this scheme degraded the longest calls in the system with a hope that those flows have higher probabilities to quite the system and leave fewer degraded connections. The authors in [15] proposed a bandwidth adaptation scheme based on per class degradation. Here, in order to admit a call, the lower possible priority class calls are degraded. Ongoing higher priority class calls are not affected by arrival of lower priority calls.There are also been policies where separate channels reserved for handoff and/or new calls. Apart from these, there are some threshold based CaC policies as well[10]. We will be basically present a parameter based CaC policy that will help us combine handoff decision and call admission.

Chapter 3

Tools & Techniques

3.1 Network Simulator 2

The ns simulator covers a very large number of applications, of protocols, of network types, of network elements and of traffic models. We call these "simulated objects". ns is an object oriented simulator, written in C++, with an OTcl interpreter as a frontend. The simulator supports a class hierarchy in C++, and a similar class hierarchy within the OTcl interpreter. The root of this hierarchy is the class TclObject. Users create new simulator objects through the interpreter.

3.1.1 Agents and Applications

In order to set up traffic flow between nodes, we must create agents and applications. The two main application used are the CBR (Constant Bit Rate) and the FTP (File Transfer Protocol) application. The Internet protocol used by FTP is TCP (Transport Control Protocol) and the one used by CBR is UDP (User Datagram Protocol).

UDP AGENTS : UDP agents are implemented in udp.cc, h. A UDP agent accepts data in variable size chunks from an application, and segments the data if needed. UDP packets also contain a monotonically increasing sequence number and an RTP timestamp. Although real UDP packets do not contain sequence numbers or timestamps, this sequence number does not incur any simulated overhead, and can be useful for tracefile analysis or for simulating UDP-based applications. The default maximum segment size (MSS) for UDP agents is 1000 byte:

Agent/UDP set packetSize_ 1000
The following commands are used to setup UDP agents in simulation scripts:

set udp0 [new Agent/UDP]
This creates an instance of the UDP agent.
$ns_ attach-agent $\langle node \rangle \langle agent \rangle$
$traffic-gen attach-agent $\langle node \rangle \langle agent \rangle$
This a class Application/Traffic/$\langle node \rangle \langle traffictype \rangle$ method which connects the traffic generator to the given $\langle node \rangle \langle agent \rangle$. For example, if we want to setup a CBR traffic flow for the udp agent, udp1, the following commands are given:
set cbr1 [new Application/Traffic/CBR]
$cbr1 attach-agent $udp1
$ns_ connect $\langle node \rangle \langle src - agent \rangle \langle node \rangle \langle dst - agent \rangle$

TCP Agents : TCP is a dynamic reliable congestion control protocol. It uses acknowledgements created by the destination to know whether packets are well received; lost packets are interpreted as congestion signals. TCP thus requires bidirectional links in order for the acknowledgements to return to the source. There are a number of variants of the TCP Protocol. Running an TCP simulation requires creating and configuring the agent, attaching an application-level data source (a traffic generator), and starting the agent and the traffic generator.

3.1.2 TraceFile

:There are a number of ways of collecting output or trace data on a simulation. Generally, trace data is either displayed directly during execution of the simulation, or (more commonly) stored in a file to be post-processed and analyzed. There are two primary but distinct types of monitoring capabilities currently supported by the simulator. The first, called traces, record each individual packet as it arrives, departs, or is dropped at a link or queue. Trace objects are configured into a simulation as nodes in the network topology, usually with a Tcl "Channel" object hooked to them, representing the destination of collected data (typically a trace file in the current directory). The other types of objects, called monitors, record counts of various interesting quantities such as packet and byte arrivals, departures, etc. Monitors can monitor counts associated with all packets, or on a perflow basis using a flow monitor. The format of trace file can be found here : `http://nsnam.isi.edu/nsnam/index.php/NS-2_Trace_Formats`.

3.2 NIST UMTS Patch

The Network Smulator 2 can be used to simulate various network and inter0networking scenarios. But the main drawback of the core ns-2 simulator is it not fully equipped with handling situations like UMTS or WiMax scenarios. Various research initiatives have been started to enable ns-2 support UMTS,Wimax etc scenarios. One such initiative is by the `nist.gov` giving us the UMTS Mobility Patch extension for ns-2.29 and another one is Enhanced UMTS Radio Access Network Extensions for ns-2(EURANE) extension for ns-2. Both were based on the almost same core source code. here we discuss the UMTS patch by NIST briefly.

3.2.1 Source & Download

The aforesaid extension package can be downloaded from `http://www.nist.gov/itl/antd/emntg/ssm_tools.cfm`.

3.2.2 Building & Installation

The building and installation instructions can be found at `http://www.nist.gov/itl/antd/emntg/upload/INSTALL_mobility.TXT`

Some useful solution to ns-2.29 and the UMTS extension installation problems/errors can be found in the following link `http://ramakrishnamundugar.blogspot.in/2012/12/ns-229-installation-in-ubuntu-1204.html`.

3.2.3 Different nodes & modules Added

Although NS-2 provides extensive support for wireless ad-hoc networks and satellite links, 3G systems such as UMTS and HSDPA were not part of the main-stream code. Several contributed modules implement UMTS functionality in NS-2 with varying levels of details. The most notable extension is EURANE and NIST UMTS patch. Some of the important new nodes added in these extensions are briefly described here:

The first and foremost, the addition of RNC - Radio Network Controller need to be mentioned. An RNC node can be created by the following piece of code :

```
$ns node-config -UmtsNodeType rnc
set RNC [$ns create-Umtsnode 0.0.0] ;# node id is 0.
$RNC set X_ 200
$RNC set Y_ 100
$RNC set Z_ 0
puts "RNC $RNC"
```

Next important node is the UMTS base station. this can be created by:

```
 $ns node-config -UmtsNodeType bs \
             -downlinkBW 384kbs \
             -downlinkTTI 10ms \
             -uplinkBW 384kbs \
             -uplinkTTI 10ms \
             -hs_downlinkTTI 2ms \
             -hs_downlinkBW 384kbs

set BS [$ns create-Umtsnode 0.0.1] ;# node id is 1
$BS set X_ 100
$BS set Y_ 100
$BS set Z_ 0
puts "BS $BS"
```

Apart from the nodes above, to support handover between heterogeneous networks, we need to have user equipment with multiple network interface cards(NIC's). To support this functionality, the concept of Multiface Node has been added in these extensions. The User Equipments(UE) and Multiface Nodes are created using the following code segment :

Creation of UMTS interface :

```
 $ns node-config -UmtsNodeType ue \
             -baseStation $BS \
             -radioNetworkController $RNC
set UMTS_UE0 [$ns create-Umtsnode 0.0.2] ;# node id is 2
$UMTS_UE0 set Y_ 50
$UMTS_UE0 set X_ 100
$UMTS_UE0 set Z_ 0
set UMTS_UE0_id [$UMTS_UE0 id]
puts "UMTS_UE0 created $UMTS_UE0_id"
```

creation WLAN Interface :

```
$ns node-config -wiredRouting OFF \
```

```
                -macTrace ON
set WLAN_UE0 [$ns node 4.0.1]
$WLAN_UE0 random-motion 0
set WLAN_UE0_id [$WLAN_UE0 id]
puts "WLAN_UE0_id $WLAN_UE0_id connet to AP0"
$WLAN_UE0 base-station [AddrParams addr2id [$AP0 node-addr]]

$WLAN_UE0 set X_ [expr 200.0]
$WLAN_UE0 set Y_ 50.0
$WLAN_UE0 set Z_ 0.0
```

creation of MultiFace Node :

```
$ns node-config  -multiIf ON
set UE0 [$ns node 5.0.0]
$UE0 set X_ 100
$UE0 set Y_ 100
$UE0 set Z_ 0
set UE0_id [$UE0 id]
puts "UE0 $UE0_id"
$ns node-config  -multiIf OFF
```

Adding the interfaces to the Multiface node :

```
# add interfaces to MultiFaceNode
$UE0 add-interface-node $WLAN_UE0
$UE0 add-interface-node $UMTS_UE0
```

3.3 Movement - Random Waypoint Mobility Model

In mobility management, the Random waypoint model is a random model for the movement of mobile users, and how their location, velocity and acceleration change over time. Mobility models are used for simulation purposes when new network protocols are evaluated. The Random waypoint model was first proposed by Johnson and Maltz. It is one of the most popular mobility models and the "benchmark" mobility model to evaluate other Mobile ad hoc network (MANET) routing protocols, because of its simplicity and wide availability. In random-based mobility simulation models, the mobile nodes move randomly and freely without restrictions. To be more specific, the destination, speed and direction are all chosen randomly and independently of other nodes. This kind of model has been used in many simulation studies. Briefly, in the RWP model:

- Each node moves along a zigzag line from one waypoint Pi to the next Pi+1.
- The waypoints are uniformly distributed over the given convex area,e.g. unit disk.
- At the start of each leg a random velocity is drawn from the velocity distribution.(in the basic case the velocity is constant 1)
- Optionally, the nodes may have so-called "thinking times" when they reach each waypoint before continuing on the next leg,where durations are independent and identically distributed random variables.

We have studied the implementation of this model in NS-2 and used it to trace the location and movement of the nodes in our simulation. The implementation details of this mobility model is as follows:

As the simulation starts, each mobile node randomly selects one location in the simulation field as the destination. It then travels towards this destination with constant velocity chosen uniformly and randomly from $[0,\mathrm{V}_{\max}]$, where the parameter $\mathrm{V}_{\max}$ is the maximum allowable velocity for every mobile node. The velocity and direction of a node are chosen independently of other nodes. Upon reaching the destination, the node stops for a duration defined by the 'pause time' parameter $\mathrm{T}_{\mathrm{pause}}$. If $\mathrm{T}_{\mathrm{pause}}=0$, this leads to continuous mobility. After this duration, it again chooses another random destination in the simulation field and moves towards it. The whole process is repeated again and again until the simulation ends. As an example, the movement trace of a node is shown in Fig.

In the Random Waypoint model, $\mathrm{V}_{\max}$ and $\mathrm{T}_{\mathrm{pause}}$ are the two key parameters that determine the mobility behavior of nodes. If the $\mathrm{V}_{\max}$ is small and the pause time $\mathrm{T}_{\mathrm{pause}}$ is long, the topology of Ad Hoc network becomes relatively stable. On the other hand, if the node moves fast (i.e.,$\mathrm{V}_{\max}$ is large) and the pause time $\mathrm{T}_{\mathrm{pause}}$ is small, the topology is expected to be highly dynamic. Varying these two parameters, especially the $\mathrm{V}_{\max}$ parameter, the Random Waypoint model can generate various mobility scenarios with different levels of nodal speed. Therefore, it seems necessary to quantify the nodal speed.

Intuitively, one such notion is average node speed. If we could assume that the pause time $T_{pause}=0$, considering that V_{max} is uniformly and randomly chosen from $[0, V_{max}]$, we can easily find that the average nodal speed is $0.5V^2_{max}$. However, in general, the pause time parameter should not be ignored. In addition, it is the relative speed of two nodes that determines whether the link between them breaks or forms, rather than their individual speeds. Thus, average node speed seems not to be the appropriate metric to represent the notion of nodal speed.[9]

3.4 Request Management - Queuing Theory

We have seen that as a system gets congested, the service delay in the system increases. A good understanding of the relationship between congestion and delay is essential for designing effective congestion control algorithms. Queuing Theory provides all the tools needed for this analysis. The following section will discuss the basic concepts and some important results.

Before we proceed further, let's understand the different components of delay in a messaging system. The total delay experienced by messages can be classified into the following categories:[7][6]

- Queuing Delay : This is the delay between the point of entry of a packet in the transmit queue to the actual point of transmission of the message.This delay depends on the load on the communication link.
- Transmission Delay : This is the delay between the transmission of first bit of the packet to the transmission of the last bit.This delay depends on the speed of the communication link.
- Propagation Delay : This is the delay between the point of transmission of the last bit of the packet to the point of reception of last bit of the packet at the other end.This delay depends on the physical characteristics of the communication link.

3.4.1 Little's Theorem

We begin our analysis of queuing systems by understanding Little's Theorem. Little's theorem states that: The average number of customers (N) can be determined from the following equation: $N = \lambda T$ Here lambda is the average customer arrival rate and T is the average service time for a customer.[7]

Proof of this theorem can be obtained from any standard textbook on queuing theory. Here we will focus on an intuitive understanding of the result. Consider the example of a restaurant where the customer arrival rate (lambda) doubles but the customers still spend the same amount of time in the restaurant (T). This will double the number of customers in the restaurant (N). By the same logic if the customer arrival rate remains the same but the customers service time doubles, this will also double the total number of customers in the restaurant.

3.4.2 Classification - M/M/1 queue

With Little's Theorem, we have developed some basic understanding of a queuing system. To further our understanding we will have to dig deeper into characteristics of a queuing system that impact its performance.

The most important characteristics of a queuing system are:

- Arrival Process
- Service Process
- No of Servers

Based on the above characteristics, queuing systems can be classified by the following convention:

A/S/n

Where A is the arrival process, S is the service process and n is the number of servers. A and S are can be any of the following:[6]

1. M(Markov) - Exponential Probability density
2. D(Deterministic) - All customers have same value
3. G(General) - Any arbitrary probability Distribution

As we will be using M/M/1 queue, we will only like to present a brief overview and important results of M/M/1 queue.

3.4.2.1 M/M/1 Queue

The notation M/M/1 indicates a queuing process with the following characteristic:[6]

- Arrival process Poisson with parameter λ.
- Service Times i.i.d, Exponential with parameter μ
- Single Server.
- Infinite Waiting Room
- N(t)- Number of customers waiting in the system at time t(state).

Stability Condition of M/M/1 queue : A queue is stable, when it does not grow to become infinite over time. The single-server queue is stable if on the average, the service time is less than the inter-arrival time, i.e. mean service time ¡ mean inter-arrival time.

Utilization Ratio : The utilization ratio of a single server M/M/1 queue is defined as the ratio between mean service time to mean inter-arrival-time. It is denoted by ρ.

$$\rho = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}} = \frac{\lambda}{\mu}$$

Some important results of M/M/1 queue has been listed below[6]:
Average Number of customers in the system :

$$N = \frac{\lambda}{\mu - \lambda} \tag{3.1}$$

Average Waiting time :

$$W = \frac{1}{\mu - \lambda} \tag{3.2}$$

Probability that the Number of customer in the system equal to a certain value :

$$P(N = n) = (\frac{\lambda}{\mu})^n \times (1 - \frac{\lambda}{\mu}) \tag{3.3}$$

3.5 Path loss Model

Path loss (or path attenuation) is the reduction in power density (attenuation) of an electromagnetic wave as it propagates through space. Path loss is a major component in the analysis and design of the link budget of a telecommunication system. This term is commonly used in wireless communications and signal propagation. Path loss may be due to many effects, such as free-space loss, refraction, diffraction, reflection, aperture-medium coupling loss, and absorption. Path loss is also influenced by terrain contours, environment (urban or rural, vegetation and foliage), propagation medium (dry or moist air), the distance between the transmitter and the receiver, and the height and location of antennas.

Causes: Path loss normally includes propagation losses caused by the natural expansion of the radio wave front in free space (which usually takes the shape of an ever-increasing sphere), absorption losses (sometimes called penetration losses), when the signal passes through media not transparent to electromagnetic waves, diffraction losses when part of the radiowave front is obstructed by an opaque obstacle, and losses caused by other phenomena.

The signal radiated by a transmitter may also travel along many and different paths to a receiver simultaneously; this effect is called multipath. Multipath waves combine at the receiver antenna, resulting in a received signal that may vary widely, depending on the distribution of the intensity and relative propagation time of the waves and bandwidth of the transmitted signal. The total power of interfering waves in a Rayleigh fading scenario vary quickly as a function of space (which is known as small scale fading). Small-scale fading refers to the rapid changes in radio signal amplitude in a short period of time or travel distance.

3.5.1 Free space path loss model

In telecommunication, free-space path loss (FSPL) is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through free space (usually air), with no obstacles nearby to cause reflection or diffraction. It does not include factors such as the gain of the antennas used at the transmitter and receiver, nor any loss associated with hardware imperfections. A discussion of these losses may be found in the article on link budget.

Free-space path loss is proportional to the square of the distance between the transmitter and receiver, and also proportional to the square of the frequency of the radio signal.

The equation for FSPL is,

$$FSPL = (\frac{4\pi d}{\lambda})^2 = (\frac{4\pi df}{c})^2 \tag{3.4}$$

where,

- λ is the signal wavelength(int metres).
- f is the signal frequency (in hertz).
- d is the distance from transmeters(in metres).
- c is the speed of light in a vacuum, 2.99792458×10^8 metres per second.

This equation is only accurate in the far field where spherical spreading can be assumed; it does not hold close to the transmitter.

3.5.2 Friis Free Space Path Loss

Apart from the formula mentioned above, the received signal power at receiver can be calculated using Friis Free Space Path loss model[8] the formula of which is given below.

The received power at a distance 'd' is given by.

$$P_{\mathrm{r}}(d) = \frac{P_{\mathrm{t}} G_{\mathrm{t}} P_{\mathrm{r}} \lambda^2}{4\pi^2 d^2 L} \tag{3.5}$$

where,

- P_t = Transmit power.
- P_r(d) = Received power at a distance 'd'.
- G_t = Transmit antenna power gain.
- G_r = Received antenna power gain.
- λ = Wave length.
- L$\geq$1 System loss factor not related to propagation.Transmission line , Filter losses, Antenna loss etc

Chapter 4

Proposed Method

4.1 Assumptions

After all the preliminary theories and approaches discussed above, let us now come to our proposed approach. We shall be giving our attention towards imperative, that means the necessary handover cases in WiFi-UMTS scenario. Our preliminary assumptions are as follows:

- There is a predefined coverage area of the WiFi
- The WiFi cell is assumed to be circular
- The AP is assumed to be located at a central position in the wifi cell.
- Mobile nodes coming out of the UMTS area into wifi area and then going out or nodes originating in the wifi coverage area and then going out.
- WiFi to WiFi handovers or UMTS to WiFi handovers, that is the horizontal handover and optional handover caes has been ignored.
- A particular WiFi area is under the coverage of more than one UMTS coverage area.
- Nodes are following a simple path loss model as explained above and also we know the emitted power of the signal from the AP and we can calculate the received power at any particular x and y.

Our proposed method can be divided into 2 phases based on the timings at which it comes into effect.

- The first phase can be termed as pre handoff
- The second phase can be termed as post handoff

In the pre handoff phase, our main objective will be to reduce the system overhead for the second phase by reducing the number of potential handovers to a certain extent. The next sections explain how do we do it.

4.2 User Remaining Request Time

Now, as we have clearly mentioned in the previous sections, handover really becomes a necessity when the user has some kind of session(call/data) going on and he is going out of a particular coverage area. In this very context, we would like to define a term user Remaining Request Time(URRT – T_r).

User Remaining Request Time(T_r) is basically the remaining time of the current session of a particular user if no handoff is required. For example, if the user is watching a video then the remaining running time of the video may be the URRT of the user or it can be the remaining call duration or internet browsing session.

Effect of URRT in handoff mainly depends on :

- cell size
- speed
- direction
- type of application running etc.

So, the effect of URRT on handoff can be characterized as a combined function of all these parameters.

Let T be the expected or estimated traveling time of the user in the present cell. There are several methods of estimating this expected traveling time in the present cell. Some of them use exponential random variables, or some estimate it by using a random direction and velocity of the mobile user. Simulations show that the ultimate outcome of these two types of estimates are almost similar if the values are taken randomly.

4.3 The Method

Phase 1 :

As per our assumption, each node is receiving power as per the path loss model. Hence, for every cell and every node, there exist a point, rather a distance from the AP at which the received power level will be too low and insufficient for any call/data session to continue. We will call that a cut off point and denote it by P_{cut}.

Apart from P_{cut}, we will define another power threshold that will be slightly higher than P_{cut}. The significance of this threshold is that, whenever any node's power level reaches this second threshold, it will basically trigger an alarm that the cut off point is nearer and/or a handover is imminent. We denote this threshold by P_{th}.

This P_{th} is actually the instance of time at which our phase 1 starts.

And it works in the following way :

1. Whenever a nodes power level reaches P_{th}, the estimated traveling time T and remaining request time T_r is found out.
2. As discussed earlier, a handover will be termed as unnecessary if the user request can be completed without opting for the handover. In our case, if we find out that $T > T_r$, then it is clear the users request will be fulfilled while he is in the present cell. Hence although P_{th} will trigger a handover alert, in this case, no actual need to transfer the request to the candidate network as the request will be in any case be taken care of before that.
3. If however $T_r > T$, then handover may be required. But in this case, we further observe that,
   (a) As the expected travel time is less than the remaining request time, the user will tend to go out of the current cell before completing the current session. This time, P_{th} will throw a positive trigger and the handoff procedure will need to be initiated. The mobile terminal will broadcast a beacon frame asking for potential candidate networks to respond and then take the decision to go to any one of them based on some criteria which will be discussed in a short while.
   (b) Here let us assume that for a particular type of cell and specific candidates, there exist a predefined handover latency that doesn't vary or varies negligibly. Let us denote the handover latency by T_d.
   (c) Also let us denote the overall lag in our candidate selection procedure be t_0. (t_0 will be clearly explained shortly.)

(d) Then, if T_r-t_0 <T_d, which basically means, the handover delay will surpass the remaining time after taking out the lag t_0 causing the session to disconnect. Hence a handover in this case,though seemingly necessary, doesn't fulfill the basic assumption of not disrupting the user session. This cases are termed as superfluous handovers and can be eliminated.

(e) In all other cases, handover is necessary.

This is the overall idea of how we are proposing to eliminate a portion of potential handover cases by classifying them as unnecessary/superfluous cases by using the user remaining request time,estimated travel time and handover delay.

Now one of the important point to be mentioned here is that the estimate of t_0 is very important.

- Early decision increases unnecessary handover probability
- Late decision may cause handover decision to be blocked and thus increase call dropping probability.
- Some methodology in the existing literature proposes advance allocation of resources for handover call. the problem with this kind of scheme is,
  1. QoS provisioning needs time, which may lead to QoS degradation and call termination at the worst case.
  2. Also, if the incoming call duration does not exceed the delay in provisioning and initializing the admission, then the allocated resource may get wasted.

To maintain a balance between them, what we propose is as follows:

We shall assume that the communication time of the beacon frames and parameter calculation time is negligible such that total lag in the process, i.e, t_0 only depends on how much the beacon frame request spends waiting in the queue at the base station.

Now if we assume for simplicity that the base stations maintain a single server queue (M/M/1) for processing requests and there are n base stations each having Poisson(λ_i) arrival rate and exponential(μ_i) service rates, i=1,2,3. . . n. Then the waiting time(average) for Base station i is $\frac{1}{\mu_i - \lambda_i}$ i=1,2. . . n.

So we need to define a measure such that all the delays are taken care of. This can be done in 2 ways:

1. We can find out the max among all the delays

2. Or we can find the average of all the delays

The first option seems more useful because by considering the max delay, we are incorporating all other delays, which means all replies from other BS will be received by then.

But the problem with this approach is that, it may take arbitrarily longer due to packet loss or some other circumstances. And due to dynamic nature of the load conditions and other parameters of the base stations, the situation may change drastically in the time between reply sent and reply received.

So, although taking average may eliminate some of the replies without waiting for them, but overall scenario may give good performance in this case.

In this way the time regarding the beacon frame transfer can be estimated.

In one approach $\mathrm{t_0}$ can be assigned as Avg$\{\frac{1}{\mu_\mathrm{i}-\lambda_\mathrm{i}}\}$over all i.

$$\mathrm{t_0} = \mathrm{avg}\ \frac{1}{\mu_\mathrm{i}-\lambda_\mathrm{i}}.$$

We can also refine this measure further by arguing that,the time lag $\mathrm{t_0}$ does not only depend on the beacon frame time, it also depends on the direction and speed in which the mobile terminal is moving towards the boundary region.

As every cell has a predefined RSS level R at which the connection is terminated. Ideally we would like to start our calculations sometimes ago at a position where the RSS will degrade below a predefined threshold, say $\mathrm{R_{th}}$.

Now, the time the mobile terminal takes to reach from $\mathrm{R_{th}}$ to R,say $\mathrm{T_b}$ plays an important part while estimating $\mathrm{t_0}$.

Let $\mathrm{t_{beacon}}$ = avg $\{\frac{1}{\mu_\mathrm{i}-\lambda_\mathrm{i}}\}$over all i.

If we take $\mathrm{t_0} = \mathrm{Max}\{\mathrm{T_b},\mathrm{t_{beacon}}\}$, then,

- if $\mathrm{t_0} = \mathrm{t_{beacon}}$, then by the above $\mathrm{t_{beacon}} > \mathrm{T_b}$ if we start that much time ago, then before fully degradation of the RSS level, we will be ready with the parameters to take a decision.
- else if $\mathrm{t_0} = \mathrm{T_b}$,then $\mathrm{t_{beacon}} < \mathrm{T_b}$, so if we start that much time ahead, then before reaching the RSS disconnection level, we will be able to get all the replies of the beacon frames.

So, in this way we try to be able to answer the when question related to the handover decision procedure a little while before the cut-off point so that we can eliminate some unnecessary/superfluous cases as well as to give ample time for the handover process to succeed. The later part, phase 2 which will incorporate the CaC scheme is discussed next.

Phase 2 :

Now the basic assumptions for this phase is as follows:

- In step 3(a) of Phase 1, when we detecting a necessary handover case, we shall be sending a beacon frame to the reachable/candidate base Station for some informations. The Informations that we will be needing in our calculation is as follows:
    1. Currectly Allocated Load
    2. Maximum Load that can be allocated
    3. Proportion of resource allocated for handoff calls
    4. load of its neighbor cell(the base station can periodically make a query to its neighbors to gather this info)
    5. Call Arrival rate
    6. Call service rate
- Secondly, we are assuming that,the time to send and receive the beacon frame is mostly dependent on the time the request spends waiting in the BS queue. Which means we are assuming the communication time to be negligible.

So, we are gathering all these information during the last part of phase 1. After the end of phase 1, if we find out that handoff is needed, we need to give the mobile user the best possible option among the available candidates so that the admission of the ongoing call is assured.This is where our phase 2, incorporating Call Admission Control(CaC) comes into effect.

We can calculate the utilization ratio[3.4.2.1] ρ of each of the base stations from Call Arrival rate and Call Service rate information.

Now we observe that, the handoff call failure rate is directly proportional to the following (for each of the base stations) :

- Proportion of allocated resource
- Proportion of allocated resource for Handoff Calls

- the data rate of the user application (the higher the data rate, higher the need of resources)
- Average wait time (3.2)
- the utilization ratio

Hence, if we define a parameter like

$$P = \frac{CL}{ML} + R_{\mathrm{HO}} + D_R + Wait_{\mathrm{avg}} + \rho \tag{4.1}$$

Where,
CL = Currently allocated load
ML = Max load
$\mathrm{R_{HO}}$ = Handoff resource allocated
D_R = Data rate
$\mathrm{Wait_{avg}}$ = average wait time in base station queue.
ρ = Utilization Ratio of the base station server.

Then, we can say that, the lower the value of this parameter, the higher the chance of the call getting admitted. Hence, if $\mathrm{P_i}$ is the value of the parameter for $\mathrm{Base\_Station_i}$ and there are say n Base stations,then, we can define the probability of the call being dropped at base station i, denoted by $\mathrm{DProb_i}$ as,

$$DProb_{\mathrm{i}} = \frac{P_{\mathrm{i}}}{\sum_{j=1}^{n} P_{\mathrm{j}}} \tag{4.2}$$

Hence,the probability of the call being admitted at base station i is,

$$AdmProb_{\mathrm{i}} = 1 - DProb_{\mathrm{i}} \tag{4.3}$$

Hence, among the candidate base stations, the one which will give the highest value of "AdmProb", we shall take the handoff decision to that base station. In this way, increasing the call admission probability apart from being able to decide whether to handover or not.

The overall procedure is being describes step by step below:

Repeat the following for each of the nodes under simulation:

(i) If the node is within the coverage area of the WLAN AP, check if the power level of the node has reached the threshold point P_{th} or not. If yes, go to step (ii) else periodically repeat step (i).

(ii) Estimate the remaining request time T_r. Also estimate the expected travel time T of the user based on its position,direction and velocity.

   a. If $T>T_r$, then mark the case as unnecessary and go to step (i).

   b. else, go to step (iii).

(iii) The mobile terminal will broadcast a beacon frame to know about the possible candidate base stations. t_0 and T_d are as explained in 4.3.

   a. if T_r-t_0 $<T_d$, then mark the case as superfluous and go to step (i).

   b. else go to step (iv).

(iv) The mobile node receives the relevant informations as explained in phase 2. Then it calculates the value of the parameter P (4.1) for each base station.

(v) The base station having the least value of that parameter is the best possible candidate for the admission of the call. The call admission probability is calculated as per equation 4.3.

We present the results in the next chapter where we will specifically show the amount of handover reductions that we have been able to achieve,the improvement over simple RSS based scheme. We shall also show that the choice given by our parameter is better most of the cases if we consider the server queue parameters only.

Chapter 5

Results

Now, in the Simulation part, we have basically used:

- NS-2.29 with UMTS patch
- Random Waypoint Mobility Model to Generate Movement Traces
- C and Matlab for generating final outcome.

For simulation purposes, we have assumed some standard values found the in literature.

- Call Arrival Rate and Call service rate at any base station has been taken such that the utilization ratio stays less than 0.9.
- We have assumed, for our simulation purpose, a single WiFi area which falls under 3 UMTS base stations.We have taken the 3 UMTS base stations with utilization ratios 0.85, 0.8 and 0.7 respectively.
- The WiFi area is assmued to be 100-120 meters.
- For path loss and received signal strength calculation, Friis Free Space Path loss model has been used.
- Movement has been generated by Random Waypoint Mobility Model
- For simplification in the simulation process, Data_Rate and some other parameter values have been scaled to the range 0-1.

We have run the simulation for 20 mobile nodes and for each node we have taken the average performance over 50 iterations.

In figure 5.1, we try to show the total number of possible handover cases in the lifetime of a given node together with the number of reductions that we have been able to do using the phase 1.

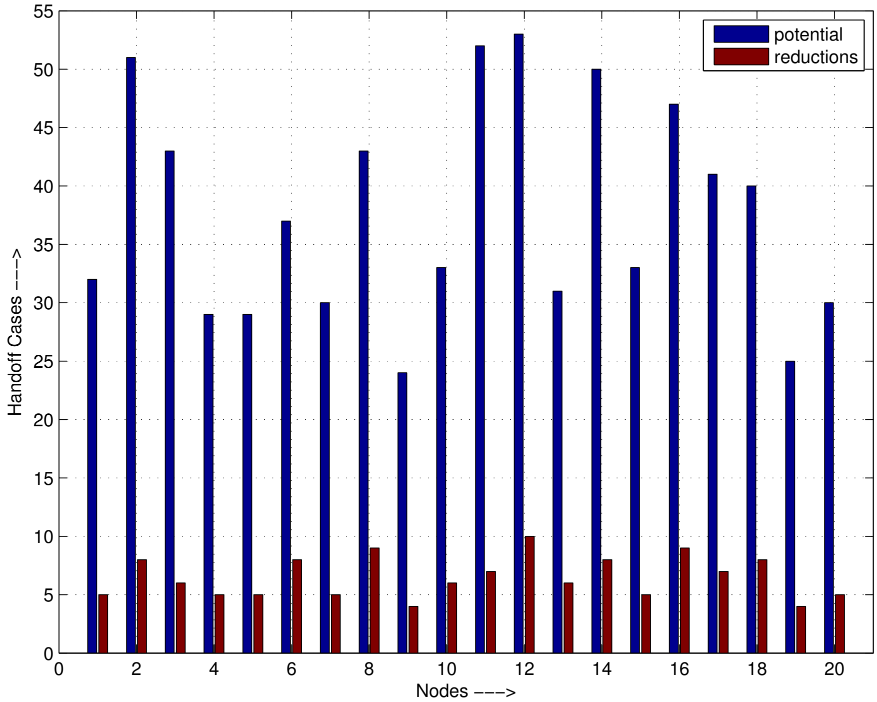

Figure 5.1: Figure showing the total number of potential handover cases and unnecessary, superfluous handover reduction for each of the 20 nodes.

In figure 5.2, the classification of the reductions(unnecessary or superfluous) have been plotted for individual node. Divided vertical bar diagram has been used showing the percentage of unnecessary and superfluous in the divisions of each vertical bar.

The next figure 5.3 plots the call admission probabilities in 2 line diagrams. One of them is for the RSS based handover case while the other one is for the parameter defined in phase 2 in our proposed method.

And in figure 5.4 we plot, for each node,total possible call admission cases averaged over 50 iterations together with the number of cases on average we are getting better result on the basis of queue at the Base station, that has been chosen by the parameter defined in phase 2.

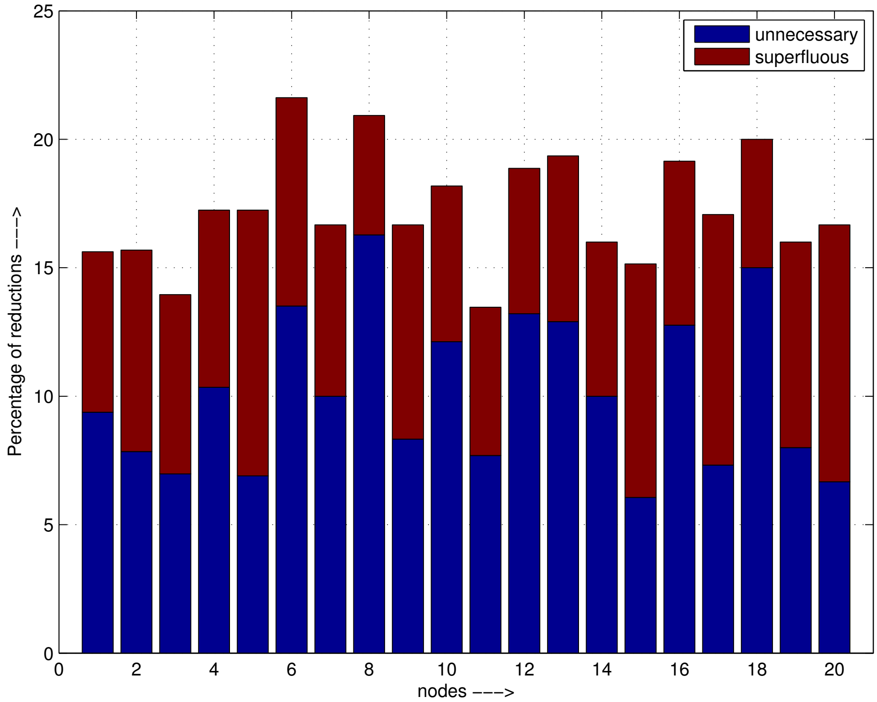

Figure 5.2: Figure showing the percentage of handovers that has been classified as unnecessary and superfluous for each node

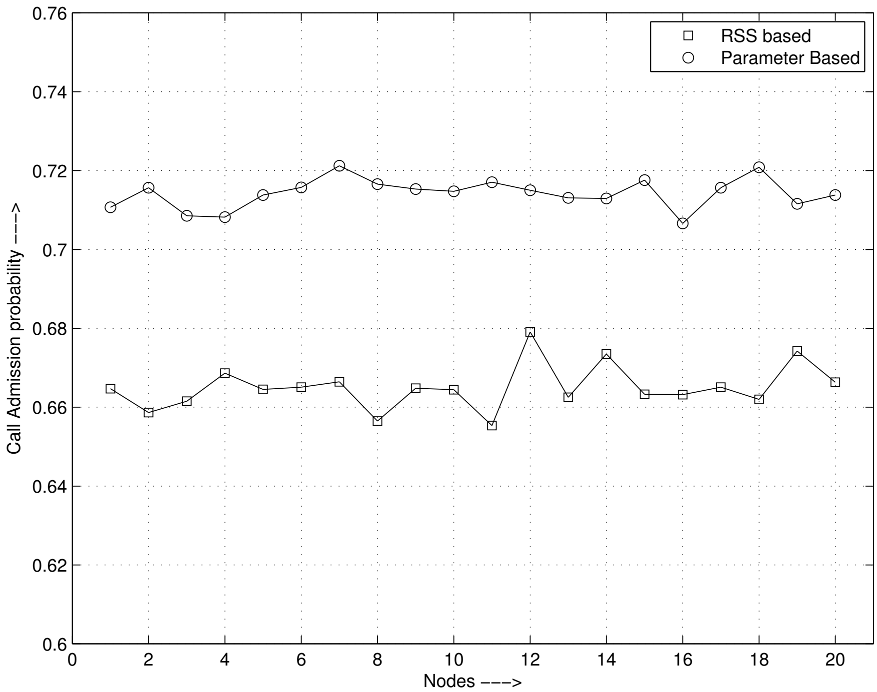

Figure 5.3: Figure comparing the Call admission probabilities of RSS based method and our proposed method

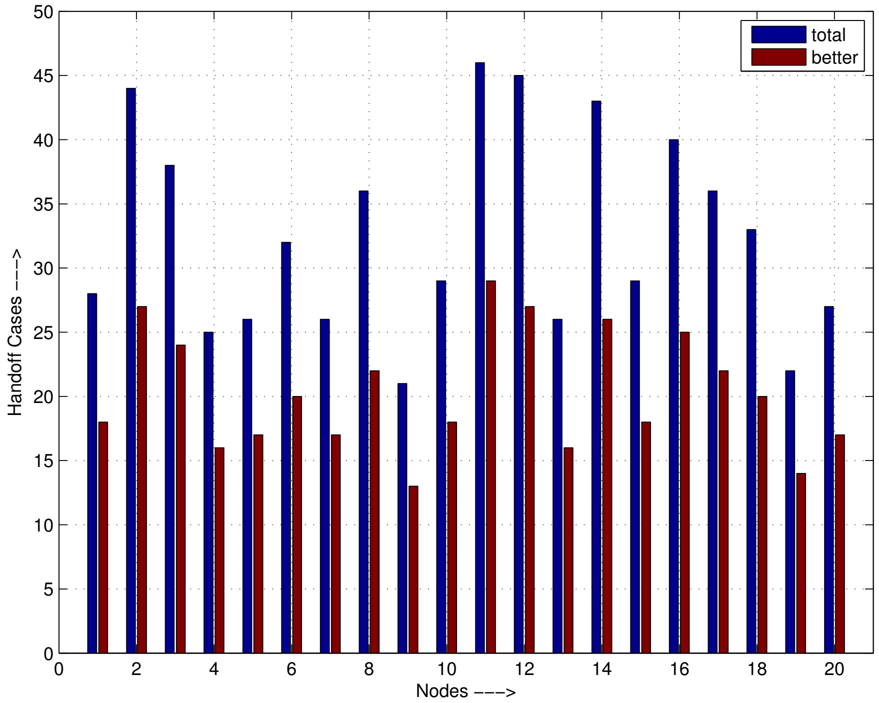

FIGURE 5.4: Grouped Vertical Bar diagram showing total call admission cases together with number of cases with better result based on the request queue at base station for each node

Bibliography

[1] H. J. Wang, R. H. Katz and J. Giese *Policy-enabled Handoffs Across Heterogeneous Wireless Networks* Proc. IEEE Wksp. Mobile Comp. Sys. and Apps., Feb 1999, pp. 51–60

[2] J. Makela, M. Ylianttila, and K. Pahlavan *Handoff Decision in Multiservice Networks* Proc. IEEE PIMRC, vol.1, Sept. 2000, pp. 655–59

[3] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Yliantillai, et al. *Handoff in Hybrid Mobile Data Networks* IEEE Personal Communications Magazine, vol. 7, no. 2, pp. 34–47, April 2000.

[4] Q. Song and A. Jamalipour *A Network Selection Mechanism for Next Generation Networks* Proc. IEEE ICC'05, Seoul, Korea, May 2005.

[5] W. Zhang *Handover Decision Using Fuzzy MADM in Heterogeneous Networks* Proc. IEEE WCNC'04, Atlanta, GA, March 2004.,

[6] Sheldon M. Ross *Introduction to Probability Models, Tenth Edition* University of Southern California Los Angeles, California

[7] `http://www.eventhelix.com/realtimemantra/congestioncontrol/queueing_theory.htm`

[8] `http://en.wikipedia.org/wiki/Friis_transmission_equation`

[9] Fan Bai and Ahmed Helmy *A SURVEY OF MOBILITY MODELS in Wireless Adhoc Networks.* University of Southern California,U.S.A

[10] Show-Shiow Tzeng *Call admission control policies in cellular wireless networks with spectrum renting* Department of Optoelectronics and Communication Engineering, National Kaohsiung Normal University, Kaohsiung 802, Taiwan

[11] S.Kokila, R.Shankar, P.Dananjayan *Analysis of Call Admission Control Schemes for WLAN Coupled to 3G Network* International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-1, Issue-6, January 2012

[12] Huan Chen et all *Guard-Channel-Based Incremental and Dynamic Optimization on Call Admission Control for Next- Generation QoS-Aware Heterogeneous Systems* IEEE Transactions on Vehicular Technology,vol.57,no.5,pp.3064–3082,September2008.

[13] W. Song, H. Jiang et all *Resource Management for QoS Support in Cellular/WLAN Interworking* IEEE Network, vol.19, no.5, pp.12-18, Sept-Oct 2005.

[14] X. G. Wang et all *An adaptive QoS framework for integrated cellular and WLAN networks*

[15] X. G. Wang et all *QoS based bandwidth management scheme in heterogeneous wireless networks* International Journal of Simulation Systems, Science and Technology, pp. 9-17, 2004.