Efficient Handover Mechanisms for Heterogeneous Networks

By

Shankar Kumar Ghosh

A thesis submitted in partial fulfillment of the requirements for the degree of *Doctor of Philosophy* in *Computer Science* at *Indian Statistical Institute*

Supervisor

Dr. Sasthi Charan Ghosh

Advanced Computing and Microelectronics Unit

Indian Statistical Institute

203 B. T. Road, Kolkata-700108

October, 2021

Dedicated

to

my beloved parents

ACKNOWLEDGEMENTS

It is with immense gratitude that I acknowledge the support of my supervisor Dr. Sasthi C. Ghosh. He gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. I deeply appreciate his valuable insight, his timely response to every query and correspondence for every submission. He has always been a source of inspiration for me. The lectures he used to deliver during course works have always helped me in understanding the research area rigorously. I am also grateful to him for giving me chances to assist in some of the courses he taught which were a very enjoyable and rewarding experience.

I would also like to register my heartfelt gratitude to all of my teachers who taught me during PhD course work and always inspired me to go beyond the limit.

I am also thankful to my lab mates for stimulating discussions, their everlasting enthusiasm, rational and free-thinking inputs.

Finally, I dedicate this work to my beloved parents whose dream it was that I wear this feather in my cap. Above all, I am thankful to God for giving me the ability, intellect and good health to carry on.

Abstract

In this thesis, some analytical frameworks have been developed to analyze the effect of different system parameters on handover performances in heterogeneous network (HetNet) and based on such frameworks, some efficient handover algorithms have been proposed. The study starts with an analytical framework to investigate the effect of resource allocation mechanisms, upper layer mobility management protocols (MMPs) and handover decision metrics on user perceived throughput. This analysis reveals that among other factors, handover decision metric plays a crucial role in determining user perceived throughput in HetNet. Subsequently, we develop two handover decision metrics for ultra dense networks (UDN) and unlicensed band communications specifically. For UDN scenario, a handover mechanism has been developed to deal with the period of time prior to handover initiation when a mobile terminal (MT) is connected to the serving cell but not getting the requested data rate. In unlicensed band communication, channel conditions may fluctuate drastically due to the interference caused by the co-existence of several networks in unlicensed band. For such a scenario, a handover mechanism has been proposed to select the optimal target network from a set of candidate networks. To provide seamless connectivity in HetNet, dual connectivity (DC) and separation between control and user planes have been emerged as promising solutions. An analytical framework has been developed to investigate the performance of DC in control-user plane split HetNet explicitly considering the data rate demands of the MTs, traffic arrival pattern and channel conditions. Finally, we analyze the handover performances of different classes of MMPs considering the effect of underlying handover execution mechanisms such as hard and semisoft. Based on this analysis, different combinations of MMPs and handover execution mechanisms have been prioritized using analytic hierarchy process. The significance of the developed analytical frameworks lie in the fact that these analyses would provide deeper insight towards service guarantee and system design in HetNet.

Publications from the content of the Thesis

In journals

- **S. K. Ghosh** and S. C. Ghosh, "Performance analysis of dual connectivity in control/user-plane split heterogeneous networks", **Computer Communications (COMCOM)**, Vol. 149, pp. 370-381, 2020. (Publisher: Elsevier, I.F. 2.816, Indexed in: Science Citation Index Expanded)
- **S. K. Ghosh** and S. C. Ghosh, "Analyzing handover performances of mobility management protocols in ultra dense networks", **Journal of Network and Systems Management (JNSM)**, Vol. 28, No. 4, pp. 1427-1452, 2020. (Publisher: Springer Nature, I.F. 2.250, Indexed in: Science Citation Index Expanded)

In conferences

- **S. K. Ghosh** and S. C. Ghosh, "Analyzing the Performance of Dual Connectivity in Control/User-plane Split Heterogeneous networks", In the proceedings of **15th Wireless On-demand Network systems and Services Conference (IEEE/IFIP WONS)**, Wengen, Switzerland, pp. 64-71, 2019.
- **S. K. Ghosh** and S. C. Ghosh, "A Predictive Handoff Mechanism for 5G Ultra Dense Networks", In the proceedings of **16th International conference of network computing and applications (IEEE NCA)**, Cambridge, MA USA, 2017.
- **S. K. Ghosh** and S. C. Ghosh, "An Analytical Framework for Throughput Analysis of Real Time Applications in All-IP Networks", In the proceedings of **31st International Conference on advanced information networking and applications (IEEE AINA)**, Taipei, Taiwan, pp. 508-515, 2017.

Manuscript under communication

- **S. K. Ghosh** and S. C. Ghosh, "A blackout aware handover mechanism for 5G ultra dense networks", submitted in **Journal of Network and Systems Management (Springer)**.

Contents

List of Figures

List of Tables

List of abbreviations

Notation	Explanation
MT	Mobile terminal
HetNet	Heterogeneous network
LTE-A	Long term evolution advanced
UMTS	Universal mobile telecommunications system
RAT	Radio access technology
WCDMA	Wideband code division multiple access
OFDMA	Orthogonal frequency division multiple access
WLAN	IEEE 802.11 wireless local area network
RSRQ	Reference signal received quality
SINR	Signal to interference plus noise ratio
DC	Dual connectivity
RLC	Radio link control
PDCP	Packet data convergence protocol
UDN	Ultra dense networks
QoS	Quality of service
MIPv6	Mobile IPv6
SIGMA	Seamless IP diversity based generalized mobility architecture
MMP	Mobility management protocol
eNB	Evolved node B
ANDSF	Access network discovery and selection function
EPC	Evolved packet core
MME	Mobility management entity
HEM	Handover execution mechanism
SGW	Serving gateway
RSRP	Reference signal received power
Hys	Hysteresis value
TTT	Time to trigger
3GPP	3rd generation partnership project
AHP	Analytic hierarchy process

Chapter 1

Introduction

The widespread use of consumer mobile terminals (MTs) such as *iPhones* and *iPads*, and extensive use of mobility supported applications such as *skype*, *facebook* and *twitter* caused mobile data traffic to be exponentially increased [1]. To cope with such high data rate demand, densification of the small cells has been emerged as one of the effective solutions by bringing the access networks closer to the MTs. Such network is referred to as ultra dense network (UDN) [2]. In UDN scenario, small cells are expected to differ in coverage ranges (10-100 meters), transmission powers (100 mW-2 W), operating frequency (licensed/unlicensed) as well as radio access technologies (RATs) such as wideband code division multiple access (WCDMA), orthogonal frequency division multiple access (OFDMA) and IEEE 802.11 wireless local area network (WLAN) [3]. Such diverse features of small cells migrate UDN towards heterogeneity. A typical heterogeneous network (HetNet) scenario is shown in Figure 1.1. Here, macrocells are providing ubiquitous coverage. Within the coverage region of the macrocells, several small cells are deployed in hotspot areas to improve system capacity. These macrocells typically belong to long term evolution advanced (LTE-A) and universal mobile telecommunications system (UMTS) standard. A HetNet scenario where all the macrocells and small cells belong to LTE-A standard is often referred to as long term evolution heterogeneous network (LTE HetNet) [1]. On the other hand, the HetNet scenario where the macrocells and small cells belong to different RATs is referred to as *all-IP* HetNet. In *all-IP* HetNet, the Internetworking protocol (IP) acts as the common interconnection medium among different RATs. Providing high data rate services to the MTs in HetNet

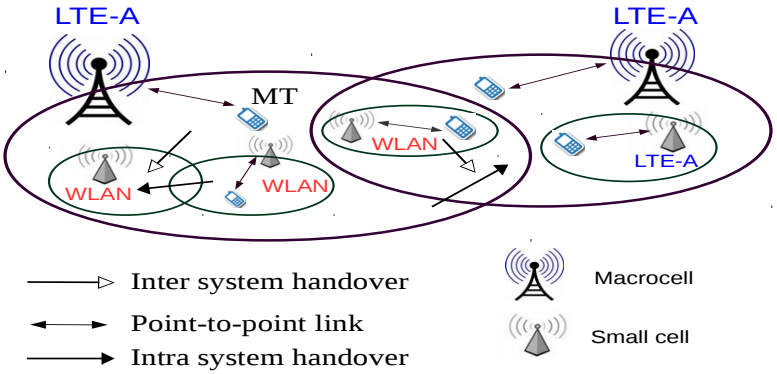


Figure 1.1: Typical HetNet scenario

scenario is one of the goals of the forthcoming fifth generation (5G) cellular systems.

While roaming across HetNet, MTs may experience frequent handovers due to the limited coverage regions of the deployed small cells. These frequent handovers cause increased control plane overhead and higher link failures. To reduce such control overhead and to ensure seamless mobility, dual connectivity (DC) and logical separation between *control plane* and *data plane* has been evolved as promising solutions [4], [5]. In control/user plane (C/U) split network architecture, macrocell evolved node Bs (eNBs) provide control coverage using a low frequency band signal and support efficient radio resource control (RRC) procedures for the MTs. Within the footprint of the macrocells, several small cells provide high data rate transmissions to the MTs over high frequency band signals. The DC technology can be used to improve per user throughput at the cell edges by utilizing resources across multiple eNBs.

Although the HetNet paradigm comes with several advantages, challenges are associated with interference management, resource allocations and seamless handovers. In particular, the handover mechanisms need to be seamless, flexible and adaptable to the variations, as changes in the traffic and deployment scenarios are expected to occur more rapidly than existing networks. In this thesis, we have developed some analytical frameworks to analyze the effect of different handover mechanisms on user throughput in various HetNet scenarios. Based on such frameworks, some new handover mechanisms have also been developed.

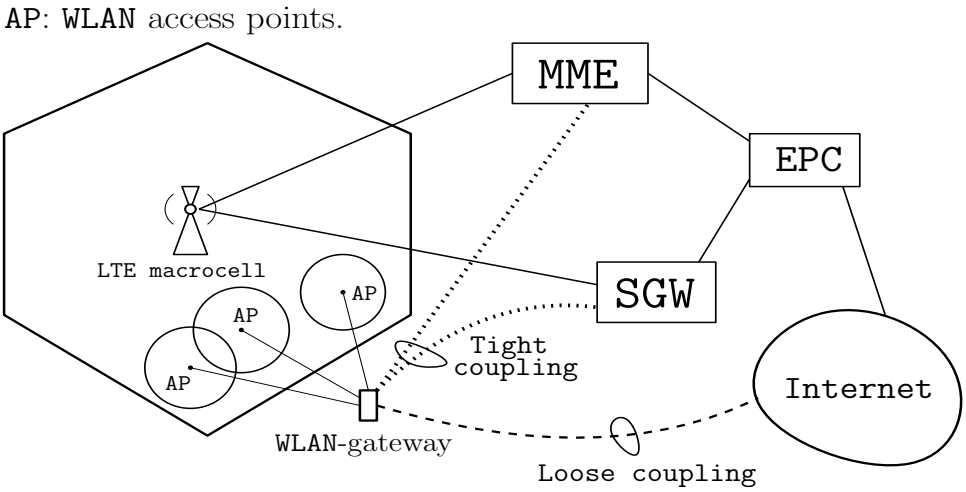


Figure 1.2: Tight coupling vs. loose coupling in *all-IP* HetNet

1.1 Background

While roaming across HetNet, MTs need to change its point of attachments from one access network to another. The mechanism of changing point of attachments from one access network to another is known as *handover*. Handovers between same RATs are called Intra system handovers (e.g., LTE−A to LTE−A handover), whereas handovers between different RATs are called Inter system handovers (e.g., LTE−A to WLAN handover). To provide *ubiquitous* connectivity across HetNet, mainly two types of coupling techniques exist in literature namely *tight coupling* and *loose coupling* [6, 7]. To illustrate the aforementioned coupling techniques, let us consider an *all-IP* HetNet consisting of one LTE macrocell and multiple WLAN access points (APs) (depicted in Figure 1.2). Here, the LTE macrocell is connected to the evolved packet core (EPC) via the mobility management entity (MME) and the serving gateway (SGW). The EPC is further connected to the Internet backbone. The WLAN APs are connected to a common gateway namely WLAN-gateway. In tight coupling, WLAN APs are connected to the EPC via WLAN-gateway. In this type of coupling, mobility management is performed by the standard mechanisms of evolved universal terrestrial radio access network (EUTRAN) which ensures lower latencies for handover and connection setup [8]. In loose coupling, WLAN APs and LTE macrocell are independently connected to the Internet. Here WLAN APs are connected to the Internet via WLAN-gateway whereas LTE macrocell is con-

nected to the Internet via EPC. In this type of coupling, the Internetworking protocol (IP) protocol acts as the common interconnecting medium and the resulting network is referred as *all-IP network* [9]. To make the handover process independent of underlying RATs in all-IP networks, several mobility management protocols (MMPs) have been proposed [10]. Examples of such MMPs include Mobile IP (operate at network layer), mobile stream control transmission protocol (operate at transport layer) and session initiation protocol (operate at application layer) [8]. In both tightly coupled and loosely coupled HetNet, handover decisions can be assisted by the access network discovery and selection function (ANDSF) which is an entity of the EPC. The goal of ANDSF is to assist an MT to discover potential 3rd generation partnership project (3GPP) (e.g., LTE) and non-3GPP (e.g., WLAN) access networks and to gather necessary information such as transmitted power, cell-traffic load and medium access control (MAC) scheduling mechanisms. In loosely coupled HetNet, MTs suffer from higher handover delay compared to tightly coupled HetNet because of the higher IP address configuration delay of MMPs. In the subsequent paragraphs, the handover mechanisms in LTE HetNet have been described. For HetNet involving multiple RATs, we describe operations of some MMPs which are frequently referred in this thesis work.

1.1.1 Handover in LTE HetNet

In LTE HetNet, both macrocell and small cells are LTE-A evolved node Bs (eNBs). However, these eNBs may differ in several aspects including coverage areas, transmission powers, access modes, radio propagation models as well as MAC access mechanisms. Typically, tight coupling is used to interconnect the eNBs in LTE HetNet, i.e., all the eNBs are connected to the EPC. The handover process in LTE-A system consists of three phases [11]. The *handover initiation phase* starts as soon as A3 event occurs, i.e., the reference signal received power (RSRP) of the neighbor eNB is higher than the RSRP of the serving eNB by a hysteresis value (Hys) for time to trigger (TTT) period of time. In *handover execution phase*, the serving network decides which neighboring access network the MT should handover based on the measurement reports send by the MTs during A3 event. Here the network selection is based on predefined *handover decision metric* such as reference signal received quality (RSRQ) and signal to interference plus noise ratio (SINR). This phase includes signaling associated with handover request,

handover request acknowledgement as well as admission control. Finally, in *handover completion phase*, resources are released from the serving cell and all active sessions are switched from serving cell to the target cell.

To deal with the throughput degradation at the cell edges, several *handover execution mechanisms* such as *hard*, *soft*, *semisoft*, *fractional soft* and *dual connectivity* (DC) have been proposed. Hard handover is a *break-before-make* mechanism where a new control link is established with the target network after releasing the old control link from the current network [12]. Consequently, the layer 2 (L2) switching delay in hard handover is very high resulting in high handover latency and packet loss.

To perform soft handover on the downlink, the 3GPP proposed a *basic* and an *optional* method [12, 13]. The idea of soft handover is to satisfy the requested data rate by combining the data rates obtained from current and target access networks at the cell edges. Clearly, soft handover has an important role in minimizing throughput degradation. To perform soft handover on the downlink, in basic method, two base stations communicate with the target MT with nearly equal power and the signal branches of current and target access network are combined using *maximal ratio combining* (MRC) technique [14]. The MRC technique works as follows. Let $\upsilon_1(t)$, $\upsilon_2(t)$, ..., $\upsilon_N(t)$ are the signals received through N diversity channels. Each of these received signals contain the originally transmitted message $m(t)$, a noise and a fading component. The MRC is a linear combining technique which maximizes the SINR of the combined output signal $\upsilon(t) = \sum_{i=1}^{N} \upsilon_i(t)$, provided (a) the noise in each channel is independent of the signal, i.e., $\upsilon_i(t) = s_i(t) + n_i(t)$ where $s_i(t)$ and $n_i(t)$ are the signal and the noise components respectively of the ith channel, (b) the signal powers are locally coherent, i.e., $s_i(t) = x_i \times m(t)$ where x_i is a positive real number that changes with time because of fading. But the rate of change is very slow compared to the instantaneous variation of $m(t)$ and (c) the noise powers are locally uncorrelated having zero means. The optional method of soft handover is based on site selection diversity transmit (SSDT) power control. In SSDT mechanism, one base station of the active set is selected for transmission to the MT. All other base stations belonging to that active set switch their power off. In LTE-A systems, soft handover is generally avoided because of its higher resource requirement and interference at the receiver end.

In [15], semisoft handover mechanism has been proposed for OFDMA systems where

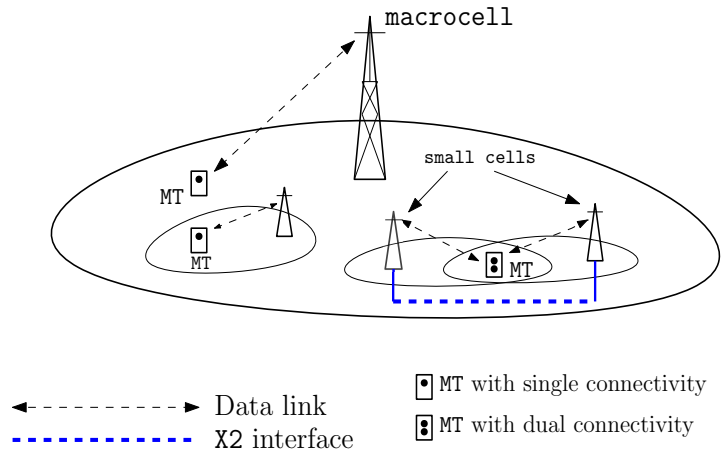


Figure 1.3: Illustration of DC mechanism

an MT can maintain control links with both the current and target base stations, but data is received only from the base station providing strongest pilot signal. To simultaneously receive control signals from several base stations using a fast fourier transform (FFT) module, the total bandwidth is divided into data and control bands. The data band is used for user data communication whereas the control band is used to communicate control information related to handover. In [16], the concept of fractional soft has been proposed. The idea of fractional handover is to receive traffic from both current and target eNBs simultaneously only for voice over IP (VoIP) service.

To deal with the high data rate demand in 5G cellular networks, recently DC technology has been proposed for LTE-A networks [5]. In principle, DC can be applied between any pair of eNBs which are connected via non-ideal X2 backhaul and operate on different frequencies (depicted in Figure 1.3). In DC mode, the concerned MT is served by two eNBs, one of which is called *master cell* and another as *secondary cell*. The DC mode enables an MT to simultaneously receive data from master cell and secondary cell which operate on different carrier frequencies. In the context of providing high data rate services, the 3C architecture of DC is of particular importance. In 3C architecture, the master cell and the secondary cell have two independent radio link control (RLC) layers and packet data convergence protocol (PDCP) layer is located only at the master cell. Here, a flow split occurs at the master cell [17] and an MT in DC mode can utilize resources across both master cell and secondary cell for the same bearer. This increases the per user throughput for a given application, which is one of the important requirement

for upcoming 5G cellular networks.

1.1.2 Handover in *all-IP* HetNet

In HetNet scenario involving multiple RATs (e.g., LTE−WLAN HetNet), typically IP protocol is used as common interconnecting medium and the resulting network is referred to as *all-IP* network. Here different RATs are independently connected to the Internet forming a loosely coupled HetNet. In this kind of HetNet, mobility is typically managed from network and transport layers by the concerned MMPs. The network layer MMPs include mobile IPv6 (MIPv6), hierarchical mobile IPv6 (HMIPv6) and fast mobile IPv6 (FMIPv6) [18, 19]. The MIPv6 protocol and its various enhancements require protocol stack modifications of the concerned MT resulting in increased complexity [20]. To address this issue, network based mobility management approaches such as proxy mobile IPv6 (PMIPv6) has been proposed [21]. In network based mobility management approach, the serving network handles the mobility management on behalf of the concerned MT. This results in reduced complexity at the concerned MT compared to that of MIPv6 based approaches. Despite such advantages, the PMIPv6 protocol often suffers from high end-to-end delay due to the redirection mechanism through the mobility anchor. Such constraint limits the performance of the delay sensitive applications. In [22], a PMIPv6 based distributed mobility management (DMM) protocol has been proposed which eliminates the necessity to redirect the packets through the mobility anchor. Here mobility related functionalities such as forwarding the ongoing sessions from one access router to another are performed by the access routers.

The MMPs operating from network layer suffer from higher end-to-end delay due to their redirection mechanism through an anchor point (e.g., home agent) in terrestrial network. To address these drawbacks, transport layer MMPs such as seamless IP diversity based generalized mobility architecture (SIGMA) [23] and mobile stream control transmission protocol (MSCTP) [24] have been proposed. The key mechanism of these protocols is to exploit the *multihoming* and *multistreaming* capacities of stream control transmission protocol (SCTP) to improve end-to-end delay and throughput for real time applications. In the next few paragraphs, functionalities of MIPv6, SIGMA and DMM are described as they are frequently referred in this thesis work.

The MIPv6 protocol [18, 19] was standardized by Internet engineering task force

(IETF) to manage mobility at IP layer. Here each MT is identified by its *home address* (HoA) assigned by *home agent* (HA), a router on an MT's home network. The HA maintains current location information for the MT and tunnels IP packets to the MT when it is away from the home network. While an MT is away from its *home network*, it is allowed to be connected with an access router of another network. This access router is commonly known as *foreign agent* (FA) and the corresponding network as *foreign network*. In foreign network, an MT is associated with a *care-of-address* (CoA) assigned by the FA. The MT always keeps the HA updated regarding its current CoA. Accordingly, a bidirectional tunnel is created in between HA and FA. To communicate with the MT, a *correspondent node* (CN) first sends the packet to the HA of the MT. The HA then tunnels the packet to the FA. Finally, the packet is delivered to the MT by the FA over the wireless link. An MT thus experiences higher end-to-end delay due to the redirection of packets though HA.

The SIGMA protocol [23] was developed to manage mobility from transport layer. The key idea of this protocol is to exploit multihoming and multistreaming capacity of stream control transmission protocol (SCTP) to improve end-to-end delay and throughput for real time applications. During handover, SIGMA acquires a new IP address from the target network and adds it to the SCTP association while the MT is still connected to the current network using its old IP address. After performing link layer handover, the old IP address become obsolete and the new one is activated. Subsequent communications are then performed using new IP address.

The DMM protocol [22] was proposed based on the architecture of PMIPv6 which operates from network layer. The DMM protocol consists of two entities namely DMM gateway (DMM-GW) and control mobility database (CMD). Users' upstream data packets are collected by the corresponding DMM-GW which are located at the access networks. Here, the DMM-GW acts as a plain access router forwarding packets to and from the Internet. The DMM-GW can also perform mobility anchoring functions. These gateways can seamlessly forward ongoing IP flows of an MT to another gateway where the MT is about to move after handover. For every MT, the CMD stores all the prefixes advertised to the MT and to which DMM-GW the MT is currently connected. By means of extended proxy binding update (PBU) and proxy binding acknowledgement (PBA) signaling, the CMD sends instructions to the DMM-GWs to recover the MT's ongoing IP flows after a

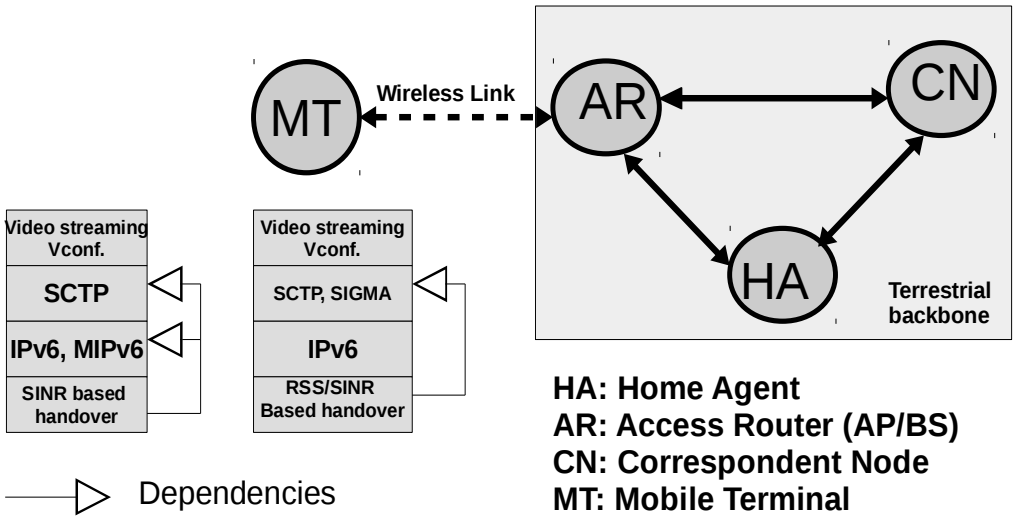


Figure 1.4: Typical network model for HetNet

handover. The DMM protocol thus avoids the redirection of packets through the network core and thereby reducing handover delay compared to PMIPv6.

The handover mechanisms used at L2 have significant impact on the performances of upper layer MMPs in *all-IP* HetNet. The L2 handover mechanism consist of three phases namely network discovery, best network determination and handover completion. In network discovery phase, an MT periodically determines the set of candidate networks. In the next phase, the best network is determined from the set of candidate networks based on a predefined metric (e.g., SINR). In handover completion phase, all active sessions are transferred from the current network to the target network.

1.2 Motivation

User throughput in *all-IP* network is influenced by the concerned MMPs considerably. This is because the packet transportation delay between two communication ends explicitly depends on the route along which the MMP delivers the packet. The MMPs operating from network layer usually suffer from higher end-to-end delay due to their redirection mechanism through an anchor point (e.g., home agent) in terrestrial network. For exam-

ple, in Figure 1.4, when MIPv6 is used for mobility management, the transmission path that a packet takes to be delivered from CN to MT is: CN $\rightarrow$ terrestrial network $\rightarrow$ HA $\rightarrow$ terrestrial network $\rightarrow$ AR $\rightarrow$ wireless link $\rightarrow$ MT. On the other hand, when SIGMA is used for mobility management of MT, the transmission path of a packet from CN to MT is: CN $\rightarrow$ terrestrial network $\rightarrow$ AR $\rightarrow$ wireless link $\rightarrow$ MT. Clearly, in SIGMA protocol, a packet need to travel a shorter path compared to that of MIPv6 as redirection through the home agent is not required. In some previous study, SIGMA has been shown to be better compared to that of MIPv6 because of its shorter transmission path [25].

Apart from the transmission path, the end-to-end delay explicitly depends on the packet loss at the wireless links. Such packet loss, in turn, depends on the handover decision metric used to select the target network. It is to be noted that the possibility of getting the requested rate from the target network exclusively depends upon the quality of service (QoS) awareness of the decision metric used by the concerned handover algorithm. For example, RSS based algorithms always select the network residing in minimum Euclidean distance as the target network. This causes load imbalance in the system and severe downfall of user perceived data rate. An MT having strict rate requirement experiences packet losses when it receives reduced data rate than the requested one. Consequently, the transport layer protocol such as SCTP invoke their congestion avoidance mechanism. This results in significant degradation of user throughput in RSS based approaches. On the other hand, SINR based algorithms consider the system load and interference level of the received signal while selecting the target network. As a result, packet loss in SINR based algorithms are much lower compared to RSS based approaches. To illustrate, let us consider that y is the expected number of transmission required to deliver a packet. After every transmission failure of the first $y-1$ transmissions, the congestion avoidance algorithm increases the retransmission time out value by a factor of α (typically 2) [26]. Finally, in y-th transmission attempt the packet got delivered to MT in $T_D^{(.)}$ time, where $T_D^{(.)}$ represents the end-to-end delay and $(.)$ is the protocol indicator. There is an upper limit m on number of possible retransmissions after which the timeout value will not be increased further. Denoting by T_o the initial timeout value,

the end-to-end packet transportation delay ($D^{(.)}$) can be computed as [26, 27]:

$$D^{(.)} = \begin{cases} \sum_{i=1}^{y-1} T_o \alpha^{i-1} + T_D^{(.)} \quad , \quad y \leq m \\ \sum_{i=1}^{m} T_o \alpha^{i-1} + T_o (y-m) \alpha^{m-1} + T_D^{(.)} \, , y > m \end{cases}$$

The SINR based algorithms perform better compared to that of RSS based algorithms as the y values for SINR based algorithms are lower compared to that of RSS based algorithms. This is because SINR measurement implicitly considers the system load while accounting interference level of the received signal. As a result, packet loss in SINR based algorithms are much lower resulting in lower y values compared to RSS. From this discussion, it is evident that the end-to-end delay incurred by the upper layer MMPs is critically influenced by the handover decision metric used at L2.

Besides the L2 handover mechanism in use, user perceived throughput is critically influenced by the MAC access mechanism of the serving network. For example, in random polling access method of WLAN, data rate obtained by individual MTs from a common access point is governed by the least data rate at which the MTs are associated with that access point [28]. On the other hand, in proportional fair access method, data rate obtained by an MT from an access point varies in proportion to the data rate at which the MT is associated with that access point [28]. In wideband code division multiple access (WCDMA) systems such as UMTS, an MT can get its requested rate if received energy per bit relative to spectral noise density $\left(\frac{E_b}{N_o}\right)$ is sufficient to get the requested rate [29]. In LTE-A systems, the user perceived throughput significantly depends on resource block (RB) allocation mechanism used by the concerned MAC access mechanism [30]. For example, in *blind equal throughput* (BET) scheduling mechanism, RBs are allocated to the MTs that have been served with lower average throughput. As a result, the MT experiencing lowest average throughput causes resource preemption. On the other hand, in proportional fair (PF) MAC access mechanism, RBs are allocated to the MTs having good channel conditions and lower past average throughput. The goal of PF is to strike a balance between requirement of fairness and spectral efficiency.

To analyze the throughput performance in HetNet, several works exist in the literature [26, 27, 31, 32]. However, these works did not considered the effect of handover

decision metric and MAC access mechanism jointly on MMPs. In [26, 27, 31], the authors have investigated only the effect of upper layer MMPs irrespective of other factors. On the other hand, in [32] and [33], authors have investigated the effect of link layer handover mechanism and proportional fair MAC access mechanism respectively on user perceived throughput, but the impact of higher layer MMPs have not been considered. It may be noted that proper combination of handover execution metric and MMPs has to be determined by the network in order to satisfy the data rate demand. In this prevailing situation, it is worthy to analyze the performances of different MMPs considering the effect of underlying handover decision metrics and MAC access mechanism. □

In UDN scenario, the MTs usually prefer to remain connected to the macrocells as the transmitting powers of macrocells are much higher compared to that of small cells [1]. While switching from a macrocell to a small cell, an MT has to defer the handover initiation until the A3 event occurs. In the mean time, throughput perceived by an MT from the serving macrocell may fall below the requested data rate because *precise estimation* of throughput is not possible only from RSRP measurements. We denote this period of time prior to handover initiation when an MT is connected to the serving cell but not getting the requested data rate as *blackout period*. To illustrate, let us consider the scenario depicted in Figure 1.5 where a macrocell is located at the point Y_1, a small cell is located at the point Y_2 and an MT is moving from macrocell to small cell through the straight line Y_1Y_2. As the MT moves away from macrocell, both the RSRP and throughput from the macrocell decreases gradually. On the other hand, the RSRP from the small cell increases steadily. The handover process from macrocell to small cell is initiated as soon as the A3 event occurs. Nonetheless, the throughput perceived from macrocell falls below the requested data rate R long before the A3 event. This intermediate period of time is referred to as blackout period. The problem of blackout period is severe in UDN scenario due to rapid fluctuation of interference level. It may be noted that the problem of satisfying data rate requests during blackout period can be addressed in the following approaches: (a) By designing effective handover initiation mechanism which can predict an upcoming blackout period and initiate handover before the perceived throughput goes below the requested data rate. (b) By designing an efficient handover decision metric which can select the target network providing the requested data rate for a long period of time and thereby avoid the blackout period. (c) By receiving data rate services

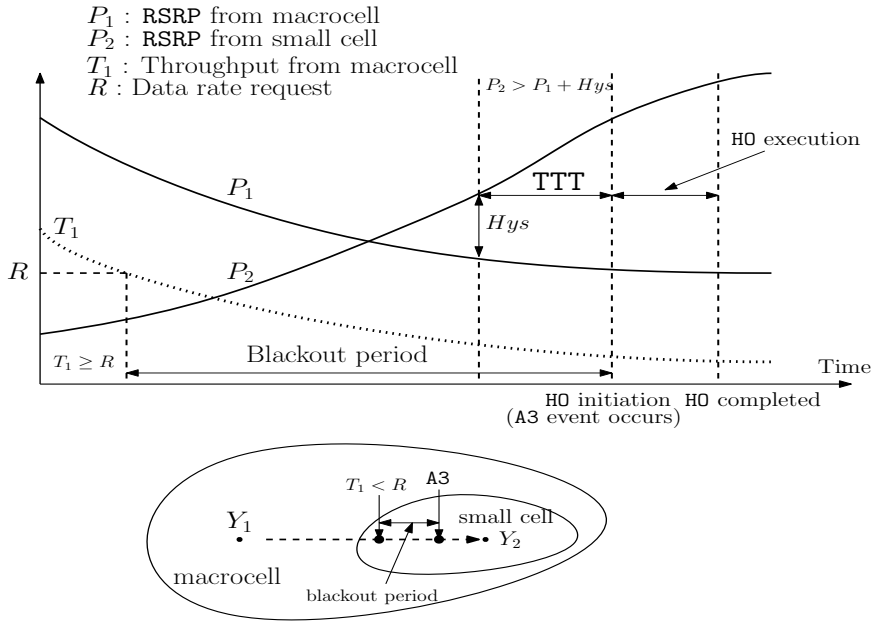


Figure 1.5: Demonstration of blackout period

simultaneously from multiple access networks to satisfy the requested data rate when a blackout period is encountered. Clearly, all of the approaches stated above call for QoS-aware handover mechanisms which can precisely estimate throughput from neighboring access networks.

The MAC access mechanisms of the serving access networks have significant impact on user throughput. Apart from MAC access mechanism, rapid fluctuation of interference level also influence the user perceived throughput in UDN scenario. The existing handover mechanisms for LTE-A systems are mainly based on received signal strength (RSS), signal to interference plus noise ration (SINR) and contextual information such as spectral efficiency, delay and jitter [34]. In [35], a reference base station efficiency (RBSE) based handover algorithm has been proposed considering the transmitted power by the base stations, traffic load as well as the users' spectral efficiency. These mechanisms do not consider the effect of MAC scheduling mechanism as well as the rapid fluctuation of interference level which are necessary for precise throughput estimation in UDN scenario. Consequently, the existing handover mechanisms can not deal with the blackout period adequately.

As the traffic load increases beyond a limit, a single access network can not provide

the requested data rate due to scarcity of RBs. In that situation, it is worthy to combine the data rates perceived from multiple access networks in order to satisfy the requested data rate. To combine the data rates from multiple access networks at the cell edges, several mechanisms such as semisoft and fractional soft have been proposed. However, these mechanisms are based on RSS measurements at the link layer. As a result, they also can not deal with the blackout period adequately. Recently, DC mechanism has been introduced to enable an MT in combining data rates from multiple access networks. However, the existing DC based algorithms [36,37] can not deal with the blackout period because of independent resource allocation at different access networks. Due to independent resource allocation, the macrocell and small cells often receive traffics from core network which is disproportionate to the capacity of the serving cells. This results in resource underutilization and packet losses which in turn causes blackout. Thus, the problem of blackout period requires a serious consideration while designing handover mechanisms for UDN scenario. □

Due to large available bandwidth and reduced cost, popularity of unlicensed band is increasing among the service providers as can be seen for LTE in unlicensed spectrum (LTE−U) and WLAN. The LTE−U has been proposed as an extension of the LTE standard so that the cellular network operators can offload some of their data traffic by accessing the unlicensed 2.4 and 5 GHz frequency band [38]. A major aspect in unlicensed band communication is the requirement to provide fair co-existence among RATs operating in the unlicensed spectrum. For example, the WLAN based network rely on contention based channel access whereas the LTE−U networks rely on schedule based channel access. As a result, starvation and other forms of unfairness may occur when the LTE−U network is deployed alongside WLAN.

While roaming across an UDN consisting of LTE−U and WLAN access networks, typically network selection is performed based on *instantaneous* measurements of the link qualities such as RSS and SINR. During the intermediate period between network selection and handover completion, channel condition of the selected target network may change drastically due to the intrinsic randomness of the radio environment. In an UDN scenario, channels operating in licensed band may exhibit intermittent characteristics due to the varying level of interference received from large number of nearby access networks. The cause of such interference is independent subcarrier allocation between

neighboring OFDMA based eNBs under universal frequency reuse (UFR) scheme [15]. In unlicensed band communication, link conditions are quite unpredictable because of the interference caused by the LTE−U and WLAN co-existence in same frequency bands (2.4 and 5 GHz) [39]. Till date there is no widely accepted co-existence mechanism to enable fair spectrum sharing between LTE−U and WLAN [39]. Due to such rapid fluctuations of link conditions, the target network which is selected based on instantaneous measurements of the link qualities may not be the *best* after the actual handover is being executed. We refer to this problem as *handover anomaly problem*. The handover anomaly problem causes higher packet losses and severe throughput degradation. Clearly, the handover anomaly problem calls for a handover mechanism which is able to predict the possibility of service guarantees from different candidate access networks after handover completion.

Designing handover mechanisms to deal with the handover anomaly problem are quite limited in the preceding literature [35, 40–42]. In [40], a context aware multi-attribute radio access technology (CMRAT) selection mechanism has been proposed based on analytic hierarchy process (AHP) utilizing contextual information. In [35], a network selection mechanism has been proposed based on a metric namely *reference base station efficiency* (RBSE) which considers the cumulative effect of transmitted powers, traffic load and user's spectral efficiency to select the target network. In [41], a context aware user-driven framework for target network selection has been proposed based on fuzzy logic. The target network selection mechanism proposed in [42] accounts network traffic load, user velocity and delay sensitivity to select the optimum target network in LTE−WLAN HetNet. It may be noted that the existing mechanisms can not deal with the rapid fluctuations of the link conditions as they are based on instantaneous measurements of various network parameters such as RSS and SINR. Moreover, the existing works are only concerned about licensed band communications. Hence, it is important to address the challenges associated with target network selection incurred by the co-existence of LTE−U and WLAN in unlicensed band. □

Although DC mechanism is effective in minimizing throughput degradation at the cell edges, this mechanism calls for resource allocations from both current and target DBSs as well. This may cause resource scarcity under highly loaded situation leading to high call drops. The amount of resource to be allocated in turn depends on several parameters

such as *data rate request*, *network traffic load*, *call arrival pattern* as well as *channel conditions*. Analyzing the performances of DC in C/U split LTE HetNet are quite limited in the preceding literature. In [43], using stochastic geometry tools, the downlink channel performance has been analyzed in terms of coverage probability and achievable rate in DC assisted HetNet. In [37], a DC assisted handover scheme has been proposed to manage handovers between macrocells in a C/U split LTE HetNet. In addition, the effect of radio link failure on handover procedures has also been analyzed while the MTs operate in DC mode. In [44] and [45], dual link handover schemes have been proposed for C/U split architecture where interruption time in U plane handover have been reduced by applying coordinated multi-point transmission and reception (CoMP) and bi-casting technology. However, these analyses [37, 43–45] do not consider the effect of realistic system parameters such as data rate requests and call arrival rates. Thus, it is important to analyze the performance of DC in C/U split HetNet. □

Frequent handovers in UDN may cause high handover packet loss and blocking rate if the handover latency is very high. High handover latency is quite unacceptable for multimedia applications such as voice over IP (VoIP) and video conferencing (Vconf). The handover latency consists of two parts: the L2 switching delay and layer 3 (L3) signaling delay. The L2 switching delay is incurred by handover execution mechanisms (HEMs) such as hard and semisoft, and the L3 signaling delay is imposed by upper layer MMPs. In hard handover, a new control link is established with the target network after releasing the old control link from current network [12]. As a result, the L2 switching delay in hard handover is very high. In contrast, in semisoft handover mechanism, an MT can maintain control links with multiple access networks simultaneously [15]. Consequently, L2 switching delay in semisoft handover is significantly less compared to that of hard handover. Generally, the network layer MMPs such as MIPv6 suffer from high handover latency due to the duplicate address detection (DAD) procedure during handover. In contrast, the SIGMA protocol exploits multihoming and multistreaming capacity of SCTP to reduce handover latency. To deal with the huge traffic volume of the forthcoming 5G UDN, PMIPv6 based DMM architecture has also been proposed [46]. In DMM, address configuration and DAD process are not part of handover latency when the handover is performed within a PMIPv6 domain. Consequently, the handover latency in DMM is expected to be lower compared to that of network layer MMPs. A typical

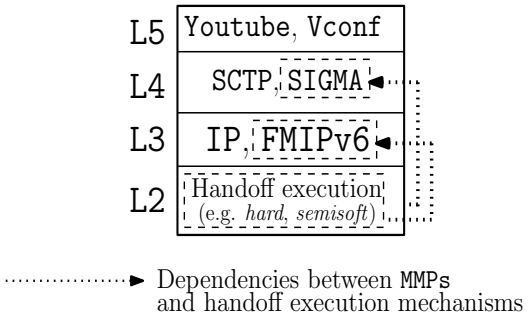


Figure 1.6: Typical protocol stack for HetNet

protocol stack showing dependencies of upper layer MMPs on HEMs are shown in Figure 1.6. The L3 signaling delay incurred by upper layer MMPs explicitly depends on the underlying HEMs used at L2. As an example, the multihoming feature enables the SIGMA protocol to maintain a pool of IP addresses (known as *association*) for future communication while only one IP address can be activated in any point of time. It may be noted that the multihoming capacity can be utilized only when semisoft handover mechanism is used at L2, i.e., the MT is capable of maintaining *multiple* control links at the cell edge. In contrast, if hard handover is used at L2 then *only one* control link is maintained by the MT with the serving access network. Consequently, the SCTP *association* contains one IP address and cannot take advantage of the multihoming capacity. Thus the IP address configuration delay becomes part of L3 handover latency, resulting in severe throughput degradation. The FMIPv6 protocol suffers from high handover latency due to the DAD procedure. However, if semisoft handover is used at the L2, then the DAD procedure can be performed using the control link associated to the target network, keeping the data communication with the current network uninterrupted. As a result, the delay associated with DAD procedure is eliminated from handover latency of FMIPv6. Similarly, in case of DMM, delay between sending the router solicitation and reception of router advertisement can be eliminated by the support of semisoft handover at L2. Despite such dependencies of signaling delay on underlying HEMs, the existing performance evaluation of MMPs [19, 21, 46, 47] do not adequately consider the effect of HEMs. In [19], various IPv6 based MMPs have been analyzed in terms of handover latency, handover blocking probability and packet loss. In [21, 46], performance evaluation of different MMPs have been carried out assuming a loosely coupled architecture. In [47],

the effect of using different handover decision metrics along with different MMPs have been analyzed on user perceived throughput for HetNet. It may be noted that proper combination of HEM and MMP has to be determined by the network in order to minimize the packet loss during handover. In this prevailing situation, it is worthy to analyze the *combined effect* of different HEMs and MMPs on user perceived throughput. □

1.3 Literature Survey

In this section, we provide a literature survey on the problems of our interest.

Throughput analysis in *all-IP* HetNet

To analyze the throughput performance in *all-IP* HetNet, several analytical models have been proposed. In [26], the effect of MMPs on different classes of applications have been analyzed in terms of transmission control protocol (TCP) level throughput and end-to-end packet transportation delay. In [31], performances of MIPv6 and SIGMA based network mobility protocols have been analyzed in terms of user perceived throughput. In [48], the performance of mobile SCTP have been analysed in terms of user perceived throughput in UMTS–WLAN HetNet. This analysis takes into account the congestion window, the round trip time, the slow-start and congestion avoidance process. In [27], authors have analyzed the throughput performance of network mobility basic support protocol (NEMO BSP) in multi-level nested network mobility environment. However, in these analyses, the effect of handover decision metric and MAC scheduling mechanism have not been considered. In [33], authors have analysed throughput considering the effect of network selection strategies of the concerned handover algorithms. Here the authors have also considered traffic and mobility characteristics of the MTs. In [32], authors have analyzed throughput in multirate WLAN explicitly considering the characteristics of PF mechanism and the load of the network. In [49], authors have analyzed spatial throughput of WLAN and LTE–U cellular networks using stochastic geometry. However, in these works, the effect of higher layer MMPs have not been considered. Thus, joint consideration of handover decision metric, higher layer MMPs and MAC scheduling mechanism is required to analyze throughput in HetNet scenario.

Challenges in ensuring service quality in UDN

To ensure better throughput performance to the MTs in HetNet scenario, a number of handover decision metrics have been proposed. A survey of such handover decision mechanisms can be found in [1, 50]. In [51, 52], either RSS or SINR has been used as fundamental decision metric. On the other hand, in [35, 42], handover decision is based on contextual information such as spectral efficiency, delay and jitter. In [51], a cross layer vertical handover has been proposed based on RSS predictions performed through polynomial regression. This mechanism enhances TCP level throughput through bandwidth estimation at MAC layer. In [52], a service adaptive multicriteria vertical handover (SAMVHO) algorithm has been proposed by jointly considering SINR, bandwidth utilization and packet loss rate. This algorithm accounts channel utilization information along with SINR from participating access networks to select the target network. It may be noted that from channel utilization information, although an MT can estimate available bandwidth, precise estimation of throughput is not possible without knowing the MAC scheduling details. In [35], a reference base station efficiency (RBSE) based network selection mechanism has been proposed for 5G UDN. The RBSE parameter takes into account the transmitted power by the base stations, traffic load as well as the users' spectral efficiency. In [42], a multi-criteria handover scheme has been proposed based on fuzzy logic which combine diverse inputs such as a users' mobility and the load of the candidate base stations. In [53], a set of cell association algorithm has been proposed with specific focus on minimizing the latency of service requested by the users in femtocell networks. Here a sequential fixing algorithm, a rounding approximation algorithm, a greedy approximation algorithm and a randomized algorithm have been proposed for cell association. It may be noted that these existing mechanisms do not consider the RB allocation policy used by different MAC scheduling mechanisms even though the RB allocation policies have a significant impact on user perceived throughput. Moreover, the rapid fluctuation of user perceived throughput in UDN scenario have also not been considered by the existing handover decision mechanisms. As a result, these metrics can not deal with the blackout period. In [54], a decision metric namely goodness has been proposed for target network selection for wideband code division multiple access (WCDMA) systems considering MAC scheduling details. However, the MAC access mechanisms of LTE−A networks are completely different from WCDMA based systems. Apart

from that the goodness based approach proposed in [54] do not consider the effect of rapid fluctuation of interference level. As a result, the metric goodness is inadequate for LTE−A based UDN scenario and therefore can not reduce the duration of blackout.

To ensure seamless communication, several data combining techniques such as soft, semisoft, fractional soft and DC have been proposed as described in section 1.1. In [55], conservative soft handover have been proposed to minimize the downfall of throughput at the cell edges. In this approach, the handover decision is based on RSS measurements and mobility management has been done by session initiation protocol (SIP). These approaches also can not address the downfall of user perceived throughput as the throughput measurement is based entirely on RSS. To enable an MT to receive data from multiple access networks, recently DC [5, 56] has been introduced for LTE−A systems. To promote the standardization of DC mechanism, several efforts have been initiated. For example, in [36] the problem of traffic splitting between macrocell and small cell in DC mechanism has been addressed considering several system parameters such as data rate requirement, maximum possible data rate request as well as capacities of backhaul links between macrocell and small cells. In [57], technical challenges associated with DC have been explored and potential solution directions have been investigated through system level simulation. These technical challenges include buffer status report calculation and logical channel prioritization. Even though the DC mechanism improves user throughput at the cell edges, it can not deal with the blackout period adequately due to independent resource allocation in adjacent OFDMA base stations. In [58], an efficient algorithm for user association has been proposed which is optimal for proportional fairness system up-to an additive constant. This algorithm can achieve significant gains in DC mode at low network load. In [59], a DC based handover protocol has been proposed that enables MTs to maintain simultaneous connections to both fourth and fifth generation cells. Here, a simulation framework has also been proposed considering the spatial dynamics of blocking events, details of medium access control, radio link control as well as transport protocols. It may be noted that the existing DC based algorithms can not measure the achievable throughput precisely due to lack of efficient throughput estimation mechanisms. To ensure the quality of services and fairness various resource allocation mechanisms have also been proposed. In [60], a rate adaptation method has been proposed considering both efficiency and fairness of the quality of experience. This

adaptation method employs a reinforcement learning mechanism to select a bit rate for each client. In [61], analytic performance models have been developed to investigate the performance of transmission control protocol (TCP) for both live and stored-media streaming. In [62], a network resource allocator system has been proposed for software defined network to enable autonomous network management based on quality of experience. However, these approaches also lack of MAC scheduling details of the candidate access networks and therefore can not deal with the blackout period.

Coexistence in unlicensed band

Although several handover mechanisms have been proposed in the recent past [8, 63], these mechanisms can not deal with the intrinsic randomness of the radio environment in UDN scenario. This is because, the existing works are based on instantaneous link quality measurements of different metrics such as RSS and SINR. To manage handovers between 3GPP and non-3GPP access networks, IEEE 802.21 media independent handover (MIH) protocol [64] has also been standardized. This protocol is also unable to manage handover mechanisms in UDN scenario because of its centralized architecture [35]. Designing handover mechanisms for UDN scenarios are quite limited in the preceding literature [35, 40–42]. In [40], a context aware multi-attribute radio access technology (CMRAT) selection mechanism has been proposed based on multiple attribute decision making (MADM) theory. The CMRAT mechanism consists of two phases. In the first phase, the CMRAT mechanism prioritizes the candidate access networks. Here the prioritization is done through context-aware AHP. In the second phase, the best network is chosen from the set of available networks based on similarities to an ideal solution. In [42], authors have proposed a network selection mechanism for a LTE-WLAN HetNet. Here target network selection is performed using fuzzy logic controller based on the contextual information such as network traffic load, user velocity and sensitivity of latency for applications gathered by the MTs. In [41], a context aware user-driven framework for target network selection has been proposed for UDN scenario. Here a fuzzy MADM methodology has been employed for selecting the target network. The existing target network selection mechanisms for UDN scenario can not deal with the handover anomaly problem. This is because, the existing works are based on instantaneous measurements of various network parameters such as RSS and SINR.

Moreover, the existing works do not address the challenges incurred by LTE−U and WLAN co-existence in unlicensed band.

Performance of DC in C/U split LTE HetNet

To promote the standardization of DC mechanism in LTE HetNet, several efforts have been initiated in recent past [36, 58, 59, 65, 66]. In these works, it has been assumed that an MT can receive data from both macrocell as well as small cells. In [65], a downlink traffic scheduling (DTS) scheme for master evolved node B (MeNB) has been proposed to split the incoming traffic between secondary evolved node B (SeNB) which will serve the MTs having DC support. The DTS scheme has been formulated as a mixed integer linear programming (MILP) problem, where the objective is to maximize the network throughput. This MILP formulation accounts the up-to-date system and network parameters such as capacities of the backhaul links, amount of downlink data that can be buffered in MeNB and SeNBs as well as data rates of bearers. In [36], the problem of traffic splitting between MeNB and SeNB for DC mechanism has been formulated as a multi-objective optimization problem (MOOP). This MOOP formulation explicitly considers several system parameters such as minimum data rate requirement for an MT, maximum possible data rate request of an MT and capacities of backhaul links between MeNB and SeNBs. Based on the characterization of MOOP, two different algorithms have been developed for throughput maximization and energy consumption minimization respectively. In [58], an efficient algorithm for user association has been proposed which is optimal for proportional fairness system upto an additive constant. This algorithm can achieve significant gains in DC mode at low network load. In [59], a DC based handover protocol has been proposed that enables MTs to maintain simultaneous connections to both fourth and fifth generation cells. Here, a simulation framework has also been proposed considering the spatial dynamics of blocking events, details of medium access control, radio link control as well as transport protocols. In [66], a network configuration for DC has been proposed based on LTE−A channel state information reference signals (CSI−RSs). This framework does not require an ideal back-haul and allows for a distributed network deployment. System-level simulations have been performed to investigate both the feasibility of the proposed network architecture as well as the performance of the cell association algorithms. In [67], the performance of DC has been

analyzed considering realistic deployment of an Europian city. Here, an opportunistic cell selection mechanism has been proposed that aims at intra-layer load balancing. In [57], technical challenges associated with DC have been explored and potential solution directions have been investigated through system level simulations. These technical challenges include buffer status report calculation and logical channel prioritization. Analyzing the performances of DC in C/U split architecture are quite limited in the preceding literature. In [43], authors have analyzed the downlink channel performance in both data and control plane for DC assisted HetNet. Here, the authors have used tools from stochastic geometry to model the coverage probability and average rate achieved by MTs in HetNet. In [37], a handover scheme exploiting the benefits of DC has been proposed to improve macrocell handover performance in C/U split HetNet architecture. In addition, the effect of radio link failure on handover procedures has also been analyzed. In [44], a network architecture for 5G C/U split heterogeneous railway wireless systems has been proposed to manage the C-plane and U-plane handovers separately. Here, the interruption time during U-plane handovers have been reduced by applying coordinated multi-point transmission and reception (CoMP). In [45], a dual-link soft handover scheme for C/U split network has been proposed for high speed railway. Here, outage probability has been reduced by applying bi-casting mechanism. In [68], an adaptive handover triggering strategy has been proposed for 5G C/U-split heterogeneous architecture to predict the received signal strength indicator and to guarantee the accuracy of the handover trigger. Performance analyses carried out in these works do not consider the effect of realistic system parameters such as data rate request, call arrival pattern and network traffic load adequately.

Interdependence of L2 and L3 handover mechanisms in UDN

Although the L3 latency incurred by upper layer MMPs depends on the underlying HEMs, the existing analyses on different MMPs do not adequately consider the combined effect. In [19], various IPv6 based MMPs have been analyzed in terms of handover latency, handover blocking probability and packet loss. In [21], analytical cost models have been developed to evaluate the performance of different IPv6 based MMPs in terms of signaling cost, packet delivery cost, tunneling cost and total cost. In [47], the combined effect of different handover decision metrics and MMPs in HetNet have been analyzed.

Here, the effect of medium access control and congestion avoidance mechanisms have been explicitly considered. In [69] and [70], the performance of network mobility basic support protocol with on-board TCP support has been compared with SIGMA based network mobility protocol for vehicular networks and satellite-terrestrial hybrid networks respectively. In [71], a PMIPv6 based DMM scheme has been proposed to improve service disruption time and packet delivery cost compared to exiting solutions. In [72], trade-offs between centralized and distributed mobility management schemes have been analyzed and a hybrid mobility management scheme has been proposed. This mobility management scheme adapts to the specific topological characteristics of the infrastructure network to significantly reduce both signaling and routing cost. These existing analyses are only concerned about L3 signaling mechanisms of upper layer MMPs without any consideration of underlying HEMs.

1.4 Scope of Thesis

In HetNet scenario, user perceived throughput depends on multiple factors such as MMP, handover decision metric, MAC scheduling mechanism used at L2 and congestion avoidance mechanism of transport layer protocols. In chapter 2, an analytical framework has been developed to analyze the effect of the above mentioned factors on user perceived throughput in *all-IP* HetNet. Using the proposed framework, the throughput performances of MIPv6 and SIGMA have been evaluated considering that the MTs have data rate request for video traffic. Extensive simulations have also been done to validate the analytical results. Our analysis reveals that the superiority of SIGMA over MIPv6 is actually *conditional* on the underlying handover decision metric.

Subsequently, in chapters 3 and 4, two handover decision metrics have been developed for HetNet scenario to deal with the blackout period and handover anomaly problem respectively. In chapter 3, a handover decision metric namely *service goodness* (SG) has been proposed considering MAC scheduling details as well as the fluctuation of interference level to deal with the blackout period. Contributions in this chapter are threefold: *Firstly,* an efficient throughput estimation algorithm has been proposed which explicitly considers the effect of MAC scheduling details as well as the rapid fluctuation of interference level in UDN scenario. The effect of MAC has been captured by jointly

considering the past average throughput, sub-carrier availability and channel condition. On the other hand, to capture the effect of fluctuating interference level leading to spiky behavior of user perceived throughput, the concept of hierarchal agglomerative clustering (HAC) has been adopted. *Secondly,* based on the proposed throughput estimation algorithm, an efficient blackout discovery mechanism and a handover decision metric namely SG have been proposed. The proposed blackout discovery mechanism can be used to predict an upcoming blackout period. On the other hand, the notion of SG of an access network represents the possibility of getting the requested rate from the concerned access network by an MT. Based on the proposed handover decision metric SG and blackout discovery mechanism, an SG based handover (SGHO) algorithm has also been proposed. System level simulations reveal that the proposed SGHO algorithm outperforms the traditional RSRP based and existing RBSE based algorithm. *Thirdly,* to deal with the blackout period at the cell edges under high load condition, a soft handover technique (SHT) has been proposed. The proposed SHT can be used with any handover algorithm to minimize the duration of blackout period. The proposed SHT also utilizes our proposed throughput estimation algorithm to measure achievable throughput from different candidate networks. The proposed SHT makes use of DC support to combine the data rates received from multiple access networks. Performance of our proposed SHT and semisoft handover have been compared based on Markov model analysis and simulations. Results obtained from analysis and simulations show that the proposed SHT significantly outperforms the semisoft handover mechanism.

In chapter 4, a predictive handover mechanism has been proposed to address the handover anomaly problem. The proposed mechanism can estimate achievable throughput values from different candidate access networks. Contributions in this chapter are as follows: *Firstly,* the target network selection problem in UDN scenario has been formulated as a *stochastic integer programming* (SIP) problem. The SIP formulation is based on the estimated throughput values that the candidate networks can provide after handover execution. Here the objective is to maximize throughput after handover execution. To deal with the handover anomaly problem, the constraint set of the SIP imposes a *lower bound* on the probability that the selected target network will be able to provide the requested data rate after handover execution. To solve the SIP, we first convert the probabilistic constraint into its equivalent deterministic one and then obtain a solution

using standard techniques involving deterministic constraints only. *Secondly,* extensive simulation has been performed to compare the proposed predictive network selection mechanism with an existing RBSE based network selection mechanism [35]. Simulation results confirm that the proposed approach significantly outperforms the RBSE based approach in terms of number of handovers, handover failure probability and goodput.

In chapter 5, an analytical framework has been proposed to compare the performances of DC with hard handover in C/U split LTE HetNet in terms of system throughput and saturation probability, i.e., the probability that the total demand for resources exceeds the total capacity of the serving cells. Contributions in this chapter can be summarized as follows: *First,* the service coverage of an MT has been modeled as the lower bound on the expected distance upto which the MT gets the requested data rate. In contrast to the existing analyses [73, 74], where service coverage regions are modeled as *circular regions* based on only received signal strength at the MTs, in this analysis, the effect of channel fading, call arrival pattern, network traffic load and data rate requests of MTs have been considered to model the service coverage region. *Then,* based on this model, we have derived the expressions for the coverage probability that an MT can be served in DC mode (P_{dc}), the probability that the MT can be served only by the current DBS (P_{bs1}), the probability that the MT can be served only by the target DBS (P_{bs2}), and the probability that the MT is residing in the *overlapping region* and can be served by both the current and target DBSs ($P_{overlap}$). These derivations explicitly consider the effect of *terminal mobility*. *Next,* based on the expression for P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$, we analyze *system throughput* as well as upper bounds on *saturation probability*. We formulate the user association problems for DC and traditional hard handover as integer linear programs (ILPs). The objectives of the formulated ILPs are to maximize system throughput. Here, the constraint set ensures that the total demand imposed by the MTs do not exceed the total capacity of the serving eNBs. Since, the demand for RBs is highly variable because of its dependence on several dynamic factors such as call arrival rate, channel fading as well as network load, we have also analyzed the upper bounds on saturation probability. Here saturation probability is defined as the probability that the total demand imposed by the MTs exceeds system capacity. *Finally*, extensive system level simulations have been performed to validate the results obtained from our proposed analytical framework. Both analyses and simulations reveal that the performance gain

of DC over traditional hard handover is actually *conditional* on underlying traffic load density and call arrival rates.

Although HEMs have significant impact on L3 handover latency, existing handover performance evaluations of MMPs do not adequately consider the effect of HEMs. In chapter 6, we have developed an analytical framework to analyze the handover performances of different classes of MMPs considering the effect of underlying HEMs in terms of handover latency, handover packet loss and handover blocking rate. We have analyzed the performances of a network layer MMP namely FMIPv6, a transport layer MMP namely SIGMA and a distributed MMP namely DMM, explicitly considering the effect of underlying HEMs such as hard and semisoft. We have considered *handover latency*, *handover packet loss* and *handover blocking rate* as performance evaluation metrics. While deriving the analytical expressions for these metrics, we consider the values of handover failure probabilities as input. Here handover failure probabilities for different HEMs have been obtained through extensive system level simulations considering the data rates of *digital cinema package* application (256 Mbps). Our analysis reveal the conditional effect of underlying HEMs and traffic load on the handover performances of FMIPv6, SIGMA and DMM in UDN scenario. Since IP address configuration delay is the principle part of handover latency and the inclusion of such delay is determined by underlying HEMs, we argue that similar observation hold for other MMPs operating from network layer, transport layer and exhibiting distributed mobility management architecture. Finally, based on these analytical results obtained from the framework, we prioritize among different combinations of MMPs and HEMs using analytic hierarchy process (AHP). This priority assignment would serve as a protocol selector in UDN scenario. Such AHP based analysis of MMPs can be generalized for any kind of application (e.g. conversational and best effort services) and HEMs by constructing the comparison matrices consistently.

1.5 Organization of Thesis

The rest of the thesis is organized as follows. In chapter 2, analytical framework for throughput analysis of real time applications in HetNet scenario has been presented. In chapter 3, the SG based handover algorithm for HetNet scenario has been described

to deal with the blackout period. In chapter 4, the predictive handover mechanism to address the handover anomaly problem has been presented. In chapter 5, we describe our proposed analytical framework to compare the performances of DC with hard handover in C/U split architecture. In chapter 6, we present the analytical framework to evaluate the joint effect of HEMs on MMPs. Finally, chapter 7 concludes the thesis.

Chapter 2

An Analytical Framework for Throughput Analysis of Real Time Applications in All-IP Networks

Guaranteeing the QoS for real time applications is one of the challenging issues in next generation all-IP networks. Analysing the user perceived throughput is essential to measure the QoS experienced by real time applications. MAC mechanism of the serving access network as well as the vertical handover (VHO) algorithm used by the MTs for target network selection have significant impact on user perceived throughput. Apart from that the upper layer MMP also influences the throughput performance considerably. The existing frameworks for throughput analysis do not adequately address the MAC specific issues. In this chapter, we have developed an analytical framework to analyze user perceived throughput considering the effect of both MAC specific details and upper layer MMPs. Using our proposed framework, we have compared the throughput performance when MIPv6 and SIGMA are used for mobility management. At L2, we have considered an RSS based and an SINR based VHO algorithm for performance evaluation. Our analysis reveals that the superiority of SIGMA over MIPv6 is actually conditional on the performance of the underlying VHO algorithm.

The rest of the chapter is organized as follows. In section 2.1, we describe the considered system model. In section 2.2, we describe our proposed analytical framework. In section 2.3, we describe performance model for MIPv6 and SIGMA. In section 2.4,

we describe the results obtained from our analytical framework. Finally, section 2.5 concludes the chapter. Important notations used in this chapter have been summarized in Table 2.1.

Table 2.1: Important notations used to analyze throughput in all-IP networks

Notation	Meaning
W^b	Carrier bandwidth of WCDMA.
C	Code division multiple access (CDMA) chip rate.
χ_0	Thermal noise power.
β	Orthogonality factor.
$\eta_{kj}(t)$	Total power received by MT j from UMTS BS k at time t.
$\phi_{ij}(t)$	Physical rate at which MT j is associated with WLAN AP i at time t.
$\xi_i(t)$	Total number of MTs associated with AP i including MT j at time t.
$\rho_{kj}(t)$	$\frac{E_b}{N_o}$ received from UMTS BS k by MT j at time t.
$\chi_{kj}(t)$	The traffic channel power received by MT j from UMTS BS k at time t.
$\psi_{ij}(t)$	The perceived throughput by MT j from WLAN AP i.
r_j^{req}	Requested data rate by MT j.
$\frac{E_b}{N_o}$	The perceived energy per bit relative to spectral noise density.
$\frac{E_b}{N_o}(r_j^{req})$	Target threshold of $\frac{E_b}{N_o}$ to get r_j^{req}.
h^a	Probability that an MT is served by a WLAN AP.
$\wp_{wl}^a$	Probability that a packet is lost due to link error of the concerned WLAN AP.
P_{loss}^A	The packet loss probability in BS−AP dual coverage areas.
P_{loss}^B	The packet loss probability in BS only coverage areas.
u^b	Probability that an MT is served by a UMTS BS.
$\wp_{wl}^b$	Probability that a packet is lost due to link error of the concerned UMTS BS.
P_{loss}	The end-to-end packet loss probability.
d_{wl}	One-way packet delivery delay at wireless link.
L^p	Processing delay at each hop.
d_{wl}^{trans}	Transmission delay at wireless link.
T_{wl}^{prop}	Propagation delay at wireless link.
T_w^{prop}	Propagation delay at wired link.
Δ_{wl}^{avg}	Time average value of user perceived data rate at the link layer.
$D^{(.)}$	Expected end-to-end delay.
$\tau^{(.)}$	The user perceived throughput from transport layer.
$\delta_w^{(.)}$	The one-way packet delivery delay at wired links.
B^w	The bandwidth of terrestrial links.
δ_d	The packet encapsulation-decapsulation delay.
$RTT^{(.)}$	The round trip timer value, $(.)$ is the indicator of upper layer MMP.
$T_D^{(.)}$	One-way packet transportation delay.

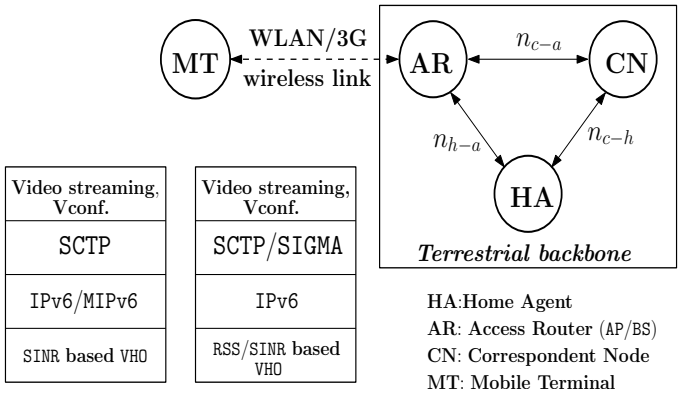


Figure 2.1: Considered network model

2.1 Considered System Model

We consider a loosely coupled UMTS−WLAN HetNet consisting of m base stations (BSs) and n access points (APs). This is due to the fact that integration of third generation (3G) technologies (e.g., UMTS) and WLAN has got much attention from research communities because of their complementary characteristics. The WLAN APs provide high data rate in small coverage regions known as hotspot areas, whereas UMTS BSs provide comparatively lower data rate over a wider coverage region. In our considered scenario, the BSs are providing ubiquitous coverage. While roaming across HetNet, an MT can be either in BS only coverage areas or in BS−AP dual coverage areas. User residence time in BS only or BS−AP dual coverage areas are assumed to be exponentially distributed with parameters ϵ_{ba} and ϵ_b. Typically these rates and branching probabilities are estimated either from the mobility traces collected from practical systems or from the simulation results [75]. In our framework, we have estimated them from the simulation results. We assume the MTs have *strict rate* requirements for video traffic. An MT is associated with the system *only if* it is getting its requested rate. Consequently, all packets communicated during the interval when an MT is not associated with the system become lost. The communication entities of the considered network model as well as possible protocol configurations for an MT are depicted in Figure 2.1. In our considered network model, we denote by n_{c-h} the average number of hops between correspondent node (CN) and home agent (HA), by n_{h-a} the average number of hops between HA and access router (AR) and by n_{c-a} the average number of hops between CN and AR. Here

mobility is being managed either by MIPv6 from network layer or by SIGMA from transport layer. At L2, either RSS based or SINR based VHO algorithm is used for target network selection. It is to be noted that the possibility of getting the requested rate from the system explicitly depends on the VHO algorithm used at L2. Accordingly, the VHO algorithm is expected to have a decisive role on both packet loss probability and transmission delay at the wireless link.

2.2 Proposed Analytical Framework

In this framework, we first model the end-to-end packet loss probability (P_{loss}) and one-way packet delivery delay at wireless link (d_{wl}) considering the effect of link layer VHO algorithm, MAC layer scheduling mechanism of the serving network and user mobility characteristics. Then, based on P_{loss} and d_{wl}, we derived the expression for end-to-end delay $D^{(.)}$. It is to be noted that $D^{(.)}$ accounts the effect of upper layer mobility management protocol. Finally, we derived user perceived throughput $\tau^{(.)}$ based on $D^{(.)}$ and P_{loss}. Here $(.)$ is the indicator of upper layer mobility management protocol such as MIPv6 or SIGMA.

2.2.1 End-to-end packet loss probability (P_{loss})

We assume that the terrestrial links are reliable and packet send over the *wired links* reach the destination MT without any retransmission trial as considered in [19]. So the end-to-end packet loss probability is entirely dependent on the loss probabilities of 3G/WLAN *wireless links*. While roaming across 3G−WLAN heterogeneous networks, an MT experiences packet losses due to the following reasons: (a) the system is not providing the requested physical rate and thereby the MT is not associated with the system. All packets communicated during this period of time become lost, and (b) the MT is getting its requested rate and thereby associated with the system, but the packet is lost due to link error occurred during transmission. We consider that an MT is *served* only if it is getting its physical rate from the system.

Let h^a be the probability that an MT is served by an AP and $\wp^a_{wl}$ be the probability that a packet is lost due to link error of the concerned AP while residing in BS−AP dual coverage areas. It is to be noted that according to the considered mobility model, the

mean residence time of an MT in BS−AP dual coverage areas is $\frac{1}{\epsilon_{ba}}$. During this time period, an MT will experience packet loss either because of not getting the requested rate (with a probability $1 - h^a$) or due to the link errors (with probability $h^a \times \wp^a_{wl}$). Hence, the packet loss probability in BS−AP dual coverage areas, P^A_{loss}, can be derived as:

$$P^A_{loss} = \frac{1}{\epsilon_{ba}} \left(1 - h^a + h^a \times \wp^a_{wl}\right) \tag{2.1}$$

Similarly, P^B_{loss}, the packet loss probability in BS only coverage areas, can be expressed as:

$$P^B_{loss} = \frac{1}{\epsilon_b} \left(1 - u^b + u^b \times \wp^b_{wl}\right) \tag{2.2}$$

where u^b and $\wp^b_{wl}$ are the probabilities that an MT is served by a BS and a packet is lost due to link error of the concerned BS while residing in BS−only coverage areas. Assuming the losses in 3G and WLAN links are independent, the end-to-end packet loss probability P_{loss} can be computed as:

$$P_{loss} \;=\; P^A_{loss} + P^B_{loss} = \frac{1}{\epsilon_{ba}} \left(1 - h^a + h^a \times \wp^a_{wl}\right) + \frac{1}{\epsilon_b} \left(1 - u^b + u^b \times \wp^b_{wl}\right) \tag{2.3}$$

It is to be noted that both h^a and u^b depends on the MAC access mechanism of the concerned AP and BS as well as on the link layer VHO algorithm run by the corresponding MT. We estimate h^a and u^b as follows.

Estimation of h^a

Assuming that APs are running *proportional fairness* [28] as MAC access mechanism, data rate perceived by MT j from AP i at time t can be computed as:

$$\psi_{ij}(t) = \frac{\phi_{ij}(t)}{\xi_i(t)} \tag{2.4}$$

Here $\phi_{ij}(t)$ is the physical rate at which the MT j is associated with AP i at time t. Note that $\phi_{ij}(t)$ belongs to a finite data rate set ϑ. In 802.11b standard, rate set is $\vartheta = \{1, 2, 5.5, 11\}$. Here $\xi_i(t)$ is the total number of MTs associated with AP i including

MT j at time t. Now, MT j will be associated with AP i if it gets the requested rate r_j^{req}. That is:

$$\psi_{ij}(t) \geqslant r_j^{req} \tag{2.5}$$

The cumulative distribution function of the random variable $\psi_{ij}(t)$ explicitly depends on the VHO algorithm run by MT j at link layer. An RSS based VHO algorithm always selects the network residing in minimum Euclidean distance as the target network. This often causes handover to a highly loaded network. Consequently, the user perceived data rate decreases. On the other hand, SINR based VHO algorithms implicitly consider network load while measuring interference level of the received signal. As a result, SINR based VHO algorithms perform better than RSS based VHO algorithms in terms of user perceived throughput [76]. For simplicity, we assume that $\psi_{ij}(t)$ is exponentially distributed with cumulative distribution function $F_a^T(t)$. Thus h^a, the probability of an MT being served in BS−AP dual coverage area, can be expressed as:

$$h^a = Pr\left(\psi_{ij}(t) \geqslant r_j^{req}\right) = 1 - F_a^T(r_j^{req}) \tag{2.6}$$

Here, the cumulative distribution function $F_a^T(t)$ has been estimated from simulation traces as has been done in [19] and [75].

Estimation of u^b

In WCDMA system such as UMTS, an MT gets its requested service if in addition to adequate pilot channel power, sufficient traffic channel power is also allocated to that MT [29]. In other words, MT j will get r_j^{req} from BS k if the perceived energy per bit relative to spectral noise density $\left(\frac{E_b}{N_o}\right)$ from BS k is greater than the target $\frac{E_b}{N_o}$ to get the requested data rate. The expression for $\rho_{kj}(t)$, the $\frac{E_b}{N_o}$ received from BS k by MT j at time t, can be written as [29]:

$$\rho_{kj}(t) = \frac{C}{r_j^{req}} \times \frac{\chi_{kj}(t)}{\chi_0 + (1-\beta)\eta_{kj(t)} + \sum_{x \neq k} \eta_{xj}(t)} \tag{2.7}$$

where $\chi_{kj}(t)$ is the traffic channel power received by MT j from BS k at time t, $\eta_{kj}(t)$ is the transmission power received by MT j from both pilot and traffic channels of BS k at time t, β is the orthogonality factor, χ_0 is the thermal noise power and C is the CDMA chip rate.

The target threshold of $\frac{E_b}{N_o}$ to get r_j^{req} can be computed as [77]:

$$\frac{E_b}{N_o}(r_j^{req}) = \frac{W^b}{r_j^{req}} \times \left(2^{\frac{r_j^{req}}{W^b}} - 1\right) \tag{2.8}$$

where W^b is the carrier bandwidth of WCDMA. Clearly, MT j will get r_j^{req} from BS k if:

$$\rho_{kj}(t) \geqslant \frac{E_b}{N_o}(r_j^{req}) \tag{2.9}$$

Similar to $\psi_{ij}(t)$, the perceived throughput by MT j from AP i, $\rho_{kj}(t)$ is also dependent on the link layer VHO as described previously. Assuming that $\rho_{kj}(t)$ is exponentially distributed with cumulative distribution function $F_b^T(t)$, the expression for u^b, the probability that an MT being served in BS-only coverage area, can be given by:

$$u^b = Pr\left(\rho_{kj}(t) \geqslant \frac{E_b}{N_o}(r_j^{req})\right) = 1 - F_b^T(r_j^{req}) \tag{2.10}$$

Similar to $F_a^T(t)$, $F_b^T(t)$ can also be estimated from simulation traces. Substituting the values of h^a (Equation 2.6) and u^b (Equation 2.10) in Equation 2.2, the expression for P_{loss} can be computed as:

$$\begin{aligned} P_{loss} &= \frac{1}{\epsilon_{ba}}\left(\wp_{wl}^a + F_a^T(r_j^{req}) - \wp_{wl}^a \times F_a^T(r_j^{req})\right) \\ &+ \frac{1}{\epsilon_b}\left(\wp_{wl}^b + F_b^T(r_j^{req}) - \wp_{wl}^b \times F_b^T(r_j^{req})\right) \end{aligned} \tag{2.11}$$

2.2.2 One-way packet delivery delay at wireless link (d_{wl})

One-way packet delivery delay at wireless link includes the processing delay (L^p), transmission delay (d_{wl}^{trans}) and the propagation delay (T_{wl}^{prop}) at the corresponding wireless link. Note that L^p and T_{wl}^{prop} can be assumed to be constants. The transmission delay can be expressed as $d_{wl}^{trans} = \frac{\theta}{\Delta_{wl}^{avg}}$ where θ and Δ_{wl}^{avg} are the segment size and time av-

erage value of user perceived data rate at the link layer respectively. Here Δ_{wl}^{avg} can be computed as:

$$\Delta_{wl}^{avg} = \frac{1}{\epsilon_{ba}} \left(h^a \times \overline{\psi_{ij}(t)} \right) + \frac{1}{\epsilon_b} \left(u^b \times \overline{\rho_{kj}(t)} \right) \tag{2.12}$$

where $\overline{\psi_{ij}(t)}$ and $\overline{\rho_{kj}(t)}$ are the time average data rate perceived by `MT` j while residing in `BS-AP dual` coverage areas and `BS-only` coverage areas respectively.

Hence the expression for one-way packet delivery delay at wireless link can be computed as:

$$\delta_{wl} = L^p + d_{wl}^{trans} + T_{wl}^{prop} = L^p + \frac{\theta}{\Delta_{wl}^{avg}} + T_{wl}^{prop} \tag{2.13}$$

After computing P_{loss} (Equation 2.11) and δ_{wl} (Equation 2.13), we derive expressions for the expected end-to-end delay $D^{(.)}$ and the user perceived throughput $\tau^{(.)}$ from transport layer considering the effect of upper layer mobility management protocol.

2.2.3 Expected end-to-end delay $\left(D^{(.)}\right)$

Expected end-to-end delay ($D^{(.)}$) is defined as the expected amount of time required to successfully deliver a packet from `CN` to `MT`. Here $D^{(.)}$ includes delay due to one-way packet transportation delay $\left(T_D^{(.)}\right)$ from `CN` to `MT` as well as *congestion avoidance* mechanism of `SCTP` caused by transmission losses. The value of $T_D^{(.)}$ includes the one-way packet delivery delay at wired links ($\delta_w^{(.)}$) as well as wireless links (δ_{wl}). Note that the delay at wired terrestrial backbone $\delta_w^{(.)}$ is protocol specific as it depends on the route along which the mobility management protocol delivers the packet. Hence, expression for $T_D^{(.)}$ can be written as:

$$T_D^{(.)} = \delta_{wl} + \delta_w^{(.)} \tag{2.14}$$

Now, to determine the delay due to congestion avoidance, we first compute the expected number of transmissions (y) required to successfully deliver a packet from `CN` to `MT` as:

$$y = \sum_{i=0}^{\infty} i \times P_{loss}^{i-1}(1 - P_{loss}) = \frac{1}{1 - P_{loss}} \tag{2.15}$$

In the expected case, after every transmission failure caused by the first $y-1$ packet losses, the congestion avoidance algorithm increases the retransmission time out value by a factor of α (typically 2) [26]. Finally, in y-th transmission attempt the packet got delivered to MT in $T_D^{(.)}$ time. There is an upper limit m on number of possible retransmissions after which the timeout value will not be increased further. Denoting by T_o the initial timeout value, the end-to-end packet transportation delay ($D^{(.)}$) can be computed following [26] and [27] as:

$$\begin{aligned} D^{(.)} &= \begin{cases} \sum_{i=1}^{y-1} T_o \alpha^{i-1} + T_D^{(.)} \quad , \quad y \leq m \\ \sum_{i=1}^{m} T_o \alpha^{i-1} + T_o (y-m) \alpha^{m-1} + T_D^{(.)} \, , y > m \end{cases} \\ &= \begin{cases} T_o \frac{\alpha^{\frac{P_{loss}}{1-P_{loss}}} - 1}{\alpha - 1} + T_D^{(.)} \, , \; \frac{1}{1-P_{loss}} \leq m \\ T_o \frac{\alpha^m - 1}{\alpha - 1} + T_o (\frac{1}{1-P_{loss}} - m) + T_D^{(.)} \, , \frac{1}{1-P_{loss}} > m \end{cases} \end{aligned} \tag{2.16}$$

2.2.4 User perceived throughput $\left(\tau^{(.)}\right)$

The user perceived throughput at *transport layer* can be measured as [31]:

$$\tau^{(.)} = \frac{\theta}{RTT^{(.)}} \times \frac{Q}{\sqrt{P_{loss}}} \tag{2.17}$$

where Q is the constant of proportionality, θ is the segment size and $RTT^{(.)}$ is the round trip timer value. Here $RTT^{(.)}$ can be estimated by doubling the value of expected end-to-end delay $D^{(.)}$. That is, $RTT^{(.)} = 2 \times D^{(.)}$.

2.3 Performance Analysis of MIPv6 and SIGMA

Using our proposed framework, in this section we analyse throughput performance when MIPv6 and SIGMA are used for mobility management. For each of these protocols, we first derive the expression for delay $\delta_w^{(.)}$ at wired terrestrial network as it depends on the route along which the MMP delivers the packet. Subsequently, we get the expressions for user throughput using Equations (2.14), (2.16) and (2.17).

When MIPv6 is used for mobility management, the transmission path that a packet takes to be delivered from CN to MT is: CN $\rightarrow$ terrestrial network $\rightarrow$ HA $\rightarrow$ terrestrial network $\rightarrow$ AR $\rightarrow$ wireless link $\rightarrow$ MT (depicted in Figure 2.1). Hence, delay at wired network when MIPv6 is used, δ_w^{mipv6} can be expressed as:

$$\delta_w^{mipv6} = \left(L^p + \frac{\theta}{B^w} + T_w^{prop} \right) \times (n_{c-h} + n_{h-a}) + \delta_d \tag{2.18}$$

where B^w is the bandwidth of terrestrial links and δ_d is the packet encapsulation and decapsulation delay. From the calculated values of δ_{wl} (Equation (2.13)) and δ_w^{mipv6} (Equation (2.18)), the one-way packet transportation delay T_D^{mipv6} when MIPv6 is in use can be computed from Equation (2.14) as: $T_D^{mipv6} = \delta_{wl} + \delta_w^{mipv6}$. Then, using the values of T_D^{mipv6} and P_{loss} (Equation (2.11)), D^{mipv6}, the expected end-to-end delay for MIPv6 can be computed from Equation (2.16). Accordingly, τ^{mipv6}, the user perceived throughput for MIPv6 can be calculated from Equation (2.17) as follows:

$$\tau^{mipv6} = \begin{cases} \frac{\theta}{2\times\left(T_o \frac{\alpha^{\frac{P_{loss}}{1-P_{loss}}}-1}{\alpha-1} + \delta_{wl} + \delta_w^{mipv6}\right)} \times \frac{Q}{\sqrt{P_{loss}}}, & \frac{1}{1-P_{loss}} \leq m \\ \frac{\theta}{2\times\left(T_o \frac{\alpha^m-1}{\alpha-1} + T_o\left(\frac{1}{1-P_{loss}} - m\right) + \delta_{wl} + \delta_w^{mipv6}\right)} \times \frac{Q}{\sqrt{P_{loss}}}, & \frac{1}{1-P_{loss}} > m \end{cases}$$

When SIGMA is used for mobility management of MT, the transmission path of a packet from CN to MT is: CN $\rightarrow$ terrestrial network $\rightarrow$ AR $\rightarrow$ wireless link $\rightarrow$ MT (depicted in Figure 2.1). Unlike MIPv6, here the packet is not redirected through HA. Hence delay in wired network when SIGMA is in use, δ_w^{sigma}, can be expressed as:

$$\delta_w^{sigma} = n_{c-a} \times \left(L^p + \frac{\theta}{B^w} + T_w^{prop} \right) \tag{2.19}$$

From the calculated values of δ_{wl} (Equation (2.13)) and δ_w^{sigma} (Equation (2.19)), T_D^{sigma}, the one-way packet delivery delay when SIGMA is in use can be computed as : $T_D^{sigma} = \delta_{wl} + \delta_w^{sigma}$. Then, using expressions for T_D^{sigma} and P_{loss} (Equation (2.11)), D^{sigma}, the expected end-to-end delay for SIGMA can be calculated from Equation (2.16). Accordingly, τ^{sigma}, the user perceived throughput for SIGMA is obtained from Equation

(2.17):

$$\tau^{sigma} = \begin{cases} \frac{\theta}{2\times\left(T_o \frac{\alpha^{\frac{P_{loss}}{1-P_{loss}}-1}}{\alpha-1}+\delta_{wl}+\delta_w^{sigma}\right)} \times \frac{Q}{\sqrt{P_{loss}}}\ , \ \frac{1}{1-P_{loss}} \leq m \\ \frac{\theta}{2\times\left(T_o \frac{\alpha^m-1}{\alpha-1}+T_o\left(\frac{1}{1-P_{loss}}-m\right)+\delta_{wl}+\delta_w^{sigma}\right)} \times \frac{Q}{\sqrt{P_{loss}}}\ , \frac{1}{1-P_{loss}} > m \end{cases}$$

2.4 Results and Discussions

To investigate the performances of different protocol configurations, we have prepared a MATLAB based simulator to carry out system level simulations. We consider that in conjunction with MIPv6 which operates from network layer for mobility management, the combined SINR based vertical soft handover (CSVSH) algorithm [76] is used at link layer for target network selection. In contrast, while evaluating the performance of SIGMA operating from transport layer, we employed both cross layer based adaptive vertical handover with predictive RSS (Pre-RSS) [51] algorithm and SINR based CSVSH algorithm for selecting the target network. The Pre-RSS [51] handoff algorithm determines the optimal target network based on RSS predictions from candidate access networks. Here the RSS predictions have been performed in two steps namely preprocessing step and prediction step. In preprocessing step, for each candidate access network, a new sequence of RSS has been generated from previous sequence of RSS by adding s' number of previous samples, where s' is a predefined integer. Next, in the prediction step, the new data sequence is used as the input data for curve fitting using polynomial regression. On the other hand, CSVSH [76] considers SINR received from different candidate networks as the fundamental decision indicator for selecting the target network. In order to make the SINR values obtained from WLAN and WCDMA networks comparable, the authors have derived a relationship between the required SINR from WLAN and received SINR from WCDMA network while offering the same downlink data rate to the MT.

2.4.1 Simulation setup

We have used the simulation environment similar to that of [78] where 7 BSs and 50 APs are deployed in a 4000×4000 square meters area. The BS coverage radius is set to 1200 meters [78]. BSs are providing *ubiquitous coverage* and the APs are placed at the boundary points of the BSs. The maximum transmitting power of BSs and APs are set to 43 dBm and 20 dBm respectively [78]. We have introduced all kind of overlapping in our simulation environment: BS $\leftrightarrow$ BS, AP $\leftrightarrow$ AP, and AP $\leftrightarrow$ BS. We consider that MTs are moving according to *smooth random* mobility model [79]. According to this model, speed behavior of an MT at time t is described by three parameters: current velocity $v(t)$ (in km/hr), current acceleration $f(t)$ (in m/s^2) and target speed $v^*(t)$. Here $v^*(t) \in \{0, 20, 60\}$ and $f(t) \in \{0, 10, 20\}$ as considered in [78]. An MT accelerates or decelerates based on current and target velocity. An MT travels with constant speed until a new target speed is decided from a predefined set. Here we have selected the target velocity from the set $\{0, 20, 60\}$ randomly.

We have considered macrocell propagation model for urban and suburban areas to calculate path loss [76]. Based on this model, pathloss (in dB) can be computed as 58.8 + $21log10(f)$ + $37.6log10(R)$ + $log(F)$, when antenna height is 15 meters. Here f is the carrier frequency (2 GHz for WCDMA and 2.4 GHz for WLAN), R is the distance in meters between the MT and the corresponding BS or AP, and $log(F)$ is the log-normal shadowing with standard deviation 10 dB. The physical rate obtained from an AP can be derived from the received signal strength indicator (RSSI) values as described in [80]. Parameter values considered in our simulation are depicted in Table 2.2.

Table 2.2: Parameter settings to evaluate MIPv6+Pre-RSS and SIGMA+CSVSH

Parameter	Value	Parameter	Value
W^b [76]	5 MHz	n_{c-h} [19]	4
n_{h-a} [19]	4	n_{c-a} [19]	6
$\wp^a_{wl}$	0.01	C [29]	3840000 Kbps
β [29]	0.5	χ_0 [76]	-99 dB
$\wp^a_{wl}$	0.1	δ_d [19]	1 ms
B^w [19]	100 Mbps	L^p [31]	1 ms
θ [31]	2 KB	T^{prop}_w [19]	0.5 ms
T^{prop}_{wl} [19]	10 ms	Q [31]	$\sqrt{\frac{3}{2}}$

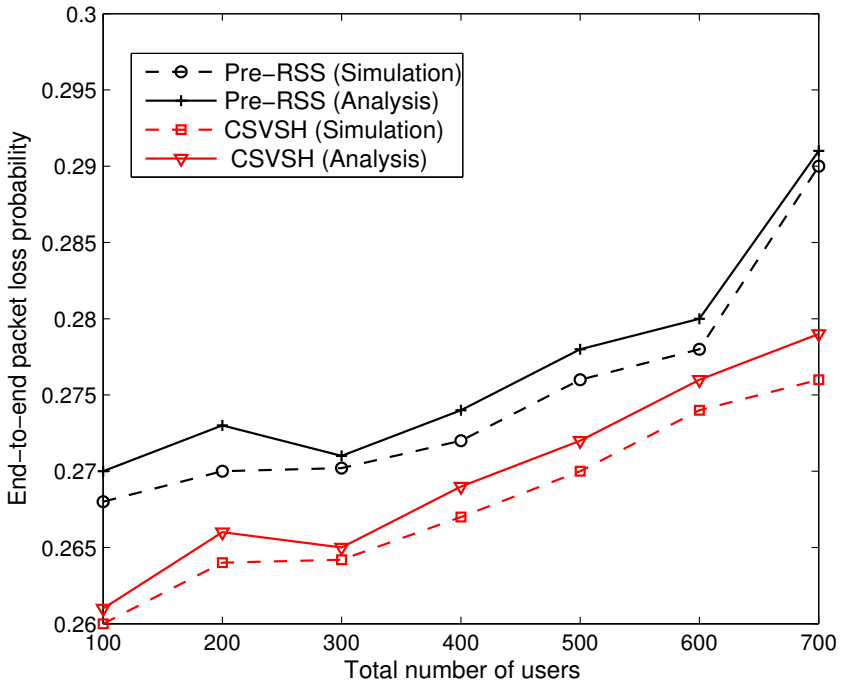


Figure 2.2: End-to-end packet loss probability vs. number of users

2.4.2 Results

Figures 2.2, 2.3 and 2.4 depict the effect of total number of users on packet loss probability, end-to-end delay and user perceived throughput respectively. The figures present both analytical and simulation results. An excellent match can be observed between the analytical and simulation results with less than 5 percent discrepancy. The users are assumed to have a typical data rate demand for video traffic (384 Kbps). We vary the traffic load in the system from 100 users to 700 users with a step of 100 users. Figure 2.2 shows the effect of traffic load on end-to-end packet loss probability. The SINR based CSVSH has been shown to outperform the RSS based Pre-RSS VHO algorithm in terms of end-to-end packet loss probability. Moreover, the packet loss probability increases with an increase in total number of users in the system. It is to be noted that the packet loss probability is governed by the losses at 3G-WLAN wireless link which in turn is dependent on both the service probabilities in BS-only coverage areas (u^b)

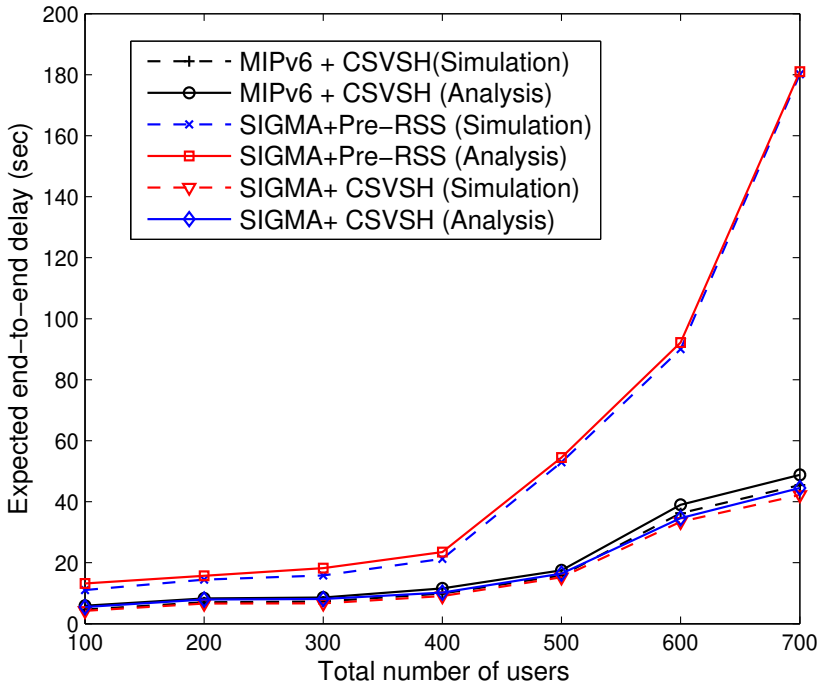


Figure 2.3: Expected end-to-end delay vs. number of users

and BS−AP dual (h^a) coverage areas. At any instant of time, an user is considered to be served if it getting its requested rate. Otherwise, the user is not associated with the system. Consequently, all packets communicated during the time interval when the user perceives a degraded data rate become lost. When RSS based Pre−RSS VHO algorithm is used for target network selection at link layer, usually the nearest BS or AP is chosen as the optimum network irrespective of its traffic load. This often leads handover to a highly loaded network. In WCDMA system, the user perceived data rate decreases with an increase in traffic load due to self interference. On the other hand, the user perceived data rate from an AP is inversely proportional to the number of users associated with that AP for the proportional fair access mechanism. Accordingly, the service probabilities decreases with an increase in system load resulting in higher packet losses. In contrast to Pre−RSS VHO algorithm, the CSVSH algorithm considers the network load of the candidate networks while measuring SINR. Due to this advanced QoS awareness, the CSVSH approach always make a handover to an access network

which is more likely to provide the requested rate to the user. Accordingly, the service probabilities in CSVSH approach is greater compared to Pre−RSS approach. As a consequence, CSVSH outperforms Pre−RSS in terms of end-to-end packet loss probability. The packet loss probability shows an increasing trend because the service probabilities decreases with an increment of system traffic load. Figure 2.3 depicts the effect of total

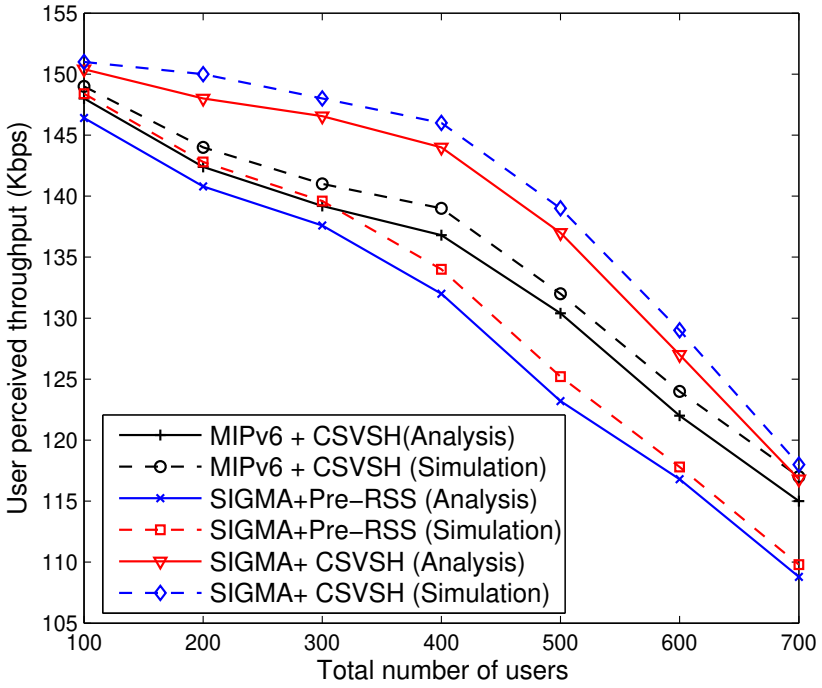


Figure 2.4: User perceived throughput vs. number of users

number of users on expected end-to-end packet transportation delay for various mobility management protocols. An MT using MIPv6 +CSVSH for mobility management suffers from greater delay at terrestrial network due to the redirection mechanism of MIPv6 through home agent, but at the same time, experiences lower delay at wireless links due to reduced packet loss probabilities caused by CSVSH. On the other hand, when SIGMA+Pre−RSS is used for mobility management, an MT experiences lower delay at terrestrial network due to end-to-end communication of SCTP, but it suffers from increased delay at wireless links due to the higher packet loss probabilities of Pre−RSS VHO algorithm. At each transmission failure SCTP protocol doubles its retransmission

timer value resulting in exponential growth of end-to-end packet transportation delay. Due to the advanced QoS awareness of link layer VHO algorithm, MIPv6+Pre-RSS outperforms SIGMA+CSVSH in terms of end-to-end delay. When, CSVSH is used with both of the mobility management protocols, SIGMA outperforms MIPv6 due to its reduced delay at terrestrial network.

Figure 2.4 depicts the effect of total number of users on user perceived throughput for different mobility management protocols. SIGMA has been shown to outperform MIPv6 when both of them are using the SINR based CSVSH algorithm at link layer for target network selection. But we get the opposite result when RSS based Pre-RSS VHO algorithm is used with SIGMA at link layer. The reason is very much similar to that described previously as user perceived throughput explicitly depends on end-to-end packet loss probability and delay.

2.5 Conclusions

We have developed an analytical framework for throughput analysis of real time applications accounting the effect of MAC access mechanism of serving network, link layer VHO algorithm, and upper layer MMP. We compared the throughput performance of video traffic application for MIPv6 and SIGMA using our proposed framework. From the analysis, we conclude that the superiority of SIGMA over MIPv6 is actually conditional on the performance of the underlying link layer VHO algorithm.

Chapter 3

A Blackout Aware Handover Mechanism for Ultra Dense Networks

While switching from a macrocell to a small cell in ultra dense networks (UDN) scenario, an MT has to defer the handover initiation until the A3 event occurs, i.e., RSRP received from serving macrocell falls below a predefined threshold. Since, RSRP can not measure the user perceived throughput precisely, the data rate perceived from the macrocell may fall below the requested data rate prior to handover. We refer to this period as *blackout period*. In UDN scenario, MTs usually prefer to remain connected to the macrocells as the transmitting powers of macrocells are much higher compared to that of small cells [1]. As a result, the MTs cannot utilize the offloading capacity of the small cells in hotspot areas predominantly. This makes the problem of blackout period more severe. The existing handover mechanisms cannot deal with the blackout period adequately as they can not deal with the rapid fluctuation of interference level. In this chapter, we propose a handover decision metric namely service goodness (SG) and a blackout discovery mechanism which consider MAC scheduling details as well as the fluctuation of interference level to deal with the blackout period. Then, we propose a SG based handover algorithm (SGHO) for UDN scenario. Simulation results confirm that our proposed SGHO outperforms the existing handover mechanisms. Furthermore, to minimize the throughput degradation at the cell edges under high load condition, we also propose a soft handover technique (SHT) which combines the data rates received from multiple access networks using DC support. Our proposed SHT can be used with

any existing handover algorithm for throughput enhancement under heavy load scenario. The superiority of SHT over semisoft handover has been established based on Markov model analysis and simulation.

Table 3.1: Important notations used to illustrate the proposed SGHO and SHT

Notation	Meaning
r_j^{req}	Requested rate of MT j.
$\gamma_{ij}(t)$	The SINR received by MT j from eNB i at t-th TTI.
$N(i,j)$	The set of all nearby active eNBs of eNB i from which MT j is experiencing co-channel interference.
$\tau_{ij}(t)$	The traffic channel power received from eNB i and MT j at t-th TTI.
$P_{col}(i)$	The sub-carrier collision probability at eNB i.
$\Gamma_{ij}(t)$	The maximum achievable throughput by MT j from eNB i at t-th TTI.
$\overline{R}_{ij}(t)$	The past average throughput experienced by the users having data rate request r_j^{req} from eNB i measured at t-th TTI.
$\nu(i)$	The number of allocated sub-carriers in eNB i.
$\Lambda(i)$	Total number of available sub-carriers in eNB i.
G_{ij}	Service goodness of eNB i with respect to MT j.
C_{ij}	Estimated throughput by MT j from eNB i.
$C_{ij}^{'}$	The rate of change of user perceived throughput at MT j from eNB i.
E_{ij}	Estimated time that MT j will get r_j^{req} from eNB i.
d^{mac}	The average MAC layer delay required to request and acquire a channel from the target eNB.
$O(j)$	The set of candidate networks constructed MT j.
$\Omega(j)$	The set of SG values of the candidate networks belonging to $O(j)$.

The rest of the chapter is organized as follows. In section 3.1, the considered system model has been described. In section 3.2, our proposed throughput estimation algorithm has been presented. Based on the proposed throughput estimation algorithm, we propose a handover decision metric namely SG and a blackout discovery mechanism in sections 3.3 and 3.4 respectively. In section 3.5, the SGHO algorithm has been proposed. In section 3.6, we present our proposed SHT. In section 3.7, we analyze the performance of our proposed SHT as well as semisoft handover mechanism based on Markov models. In section 3.8, we describe simulation results. Finally, section 3.9 concludes the chapter. Important notations used in this chapter have been summarized in Table 3.1.

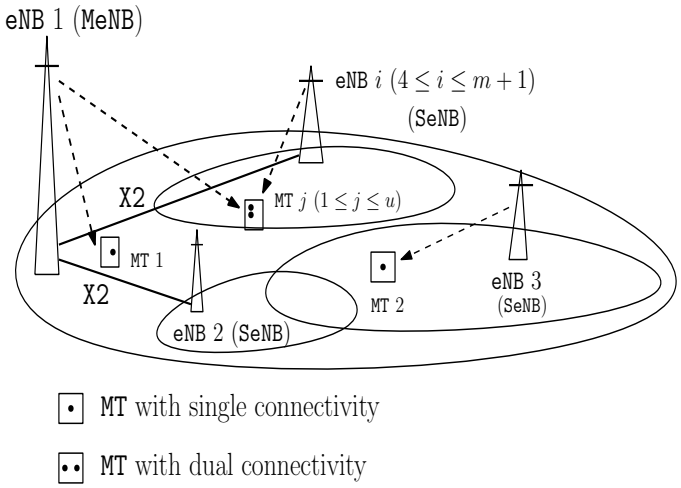


Figure 3.1: An instance of HetNet consisting of n number of MeNBs and m number of SeNBs

3.1 Considered System Model

We consider an LTE-A HetNet consisting of $m + n$ number of evolved node Bs (eNBs) including n number of macro cell evolved node Bs (MeNBs) and m number of small cell evolved node Bs (SeNBs) (depicted in Figure 3.1). Without loss of generality, we represent by eNB 1 to eNB n as MeNBs and by eNB $n + 1$ to eNB $n + m$ as SeNBs. Here the MeNBs are providing ubiquitous coverage. As considered in 3rd generation partnership project (3GPP) release 12 [5], the distance between MeNBs has been considered to be 500 meters. Within the coverage region of the MeNBs, the SeNBs are randomly deployed in condensed clusters. Here the minimum distance between SeNBs has been set to 20 meters. We assume that the MeNBs and SeNBs are connected through non-ideal X2 interface having limited capacity. We consider that u number of MTs are assumed to be *uniformly* distributed and roaming according to *smooth random waypoint* mobility model [79]. We assume that each MT j $(1 \leq j \leq u)$ has *strict rate* requirement r_j^{req}. In the considered scenario, MTs are assumed to be DC supported. In our considered DC mechanism, traffic arrived from core network are split at MeNBs and a fraction of the traffic is forwarded to the connected SeNBs over the X2 interface. Here MT j is benefited in DC mode if the aggregated data rate from multiple access networks is greater than r_j^{req}. In this work, we concentrate on downlink traffic as they require higher

bandwidth than that of uplink traffic [76]. We also consider that the eNBs are using PF as MAC access mechanism. Assuming *intra-cell* interference and background noise are negligible compared to *inter-cell* interference in OFDMA systems, $\gamma_{ij}(t)$, the SINR received by MT j from eNB i $(1 \leq i \leq m+n)$ at t-th transmission time interval (TTI) has been computed as [15]:

$$\gamma_{ij}(t) = \frac{\tau_{ij}(t)}{P_{col}(i) \sum\limits_{v \in N(i,j)} \tau_{vj}(t)} \tag{3.1}$$

where $N(i,j)$ is the set of all nearby active eNBs of eNB i from which MT j is experiencing co-channel interference, $\tau_{ij}(t)$ is the traffic channel power received from eNB i by MT j at time t and $P_{col}(i)$ is the sub-carrier collision probability at eNB i. Here $P_{col}(i)$ has been computed as:

$$P_{col}(i) \quad = \quad \frac{\nu(i)}{\Lambda(i)} \tag{3.2}$$

where $\nu(i)$ and $\Lambda(i)$ represent the number of allocated sub-carriers in eNB i and total number of available sub-carriers in eNB i respectively. Based on $\gamma_{ij}(t)$, the maximum achievable throughput $\Gamma_{ij}(t)$ by MT j from eNB i at t-th TTI has been computed as [59]:

$$\Gamma_{ij}(t) = \frac{S_r S_e}{T_s} \log_2 \left(1 + \gamma_{ij}(t)\right) \tag{3.3}$$

where S_r is the number of sub-carriers in a RB, S_e is the number of effective OFDMA symbols per LTE-A frame and T_s is the frame length. It may be noted that the $\Gamma_{ij}(t)$ samples capture the varying channel conditions at different TTIs. In the next section, we propose a throughput estimation algorithm which utilizes the $\Gamma_{ij}(t)$ samples.

3.2 Proposed Throughput Estimation Algorithm

In this section, to estimate throughput from nearby access networks, we propose a throughput estimation algorithm which explicitly considers the effect of MAC scheduling mechanism as well as *spiky behavior* of user perceived data rate in UDN scenario. Our

proposed throughput estimation algorithm consists of three phases. In *sample vector preparation* phase, our proposed algorithm jointly considers past average throughput, sub-carrier availability and channel conditions to capture the effect of MAC scheduling mechanisms. In *cluster formation* phase, to capture the effect of the spiky behavior of user perceived data rates caused by varying channel conditions at different TTIs, we have employed the hierarchical agglomerative clustering (HAC) algorithm [81]. The HAC algorithm categorizes the data rate samples into three clusters containing *low*, *moderate* and *high* data rate samples. Finally, in *throughput computation* phase, the estimated throughput is computed as an weighted average of the data rate samples belonging to different clusters. Here, weight assignments are based on cluster size.

Our proposed throughput estimation algorithm can be implemented through access network discovery and selection function (ANDSF) [35]. The role of the ANDSF is to assist the MTs in discovering potential 3GPP/non-3GPP access networks and to be associated with them. In our case, the ANDSF assists the MT in collecting necessary measurements such as sub-carrier availability and past average throughput values of users having similar data rate services in nearby access networks, and in computing the estimated throughput based on the proposed HAC based algorithm. In the subsequent subsections, we describe the steps of our proposed algorithm in detail.

Sample vector preparation

For each MT j and eNB i, the serving access network of MT j maintains an one dimensional array $V_{ij}[\,]$ of length l, where l is the number of RSRP samples retained by eNBs for network discovery purpose. Here $V_{ij}[k]$ contains the throughput value for MT j from eNB i estimated at k-th TTI $(1 \leq k \leq l)$. Based on the channel quality information (CQI) of eNB i reported by MT j, our algorithm computes $V_{ij}[k]$ as follows:

$$V_{ij}[k] = P_{col}(i)\overline{R}_{ij}(k) + (1 - P_{col}(i))\,\Gamma_{ij}(k) \tag{3.4}$$

Here $\overline{R}_{ij}(k)$ denotes the past average throughput experienced by the users having data rate request r_j^{req} from eNB i measured at k-th TTI. Typically, such values remain available to the eNBs for scheduling purposes [30]. It may be noted that the value of $\overline{R}_{ij}(k)$ explicitly depends on the MAC access mechanism used by eNB i. The value of $\Gamma_{ij}(k)$

has been computed using equation (3.3). It may be noted that $\Gamma_{ij}(k)$ represents the theoretical upper bound on $\overline{R}_{ij}(k)$. Unit of both $\Gamma_{ij}(k)$ and $\overline{R}_{ij}(k)$ is *bits/sec*. The value of $P_{col}(i)$ has been computed using equation (5.5). A higher $P_{col}(i)$ indicates higher number of sub-carrier allocations at eNB i which in turn indicates a lower possibility of MT j being scheduled by eNB i and vice-versa.

While computing $V_{ij}[k]$, $P_{col}(i)$ has been used as a weighting factor. When $P_{col}(i)$ is high, i.e., the probability of MT j being scheduled is low, then the past average throughput $\overline{R}_{ij}(k)$ gets higher priority. On the other hand, when $P_{col}(i)$ is low, i.e., the probability that MT j will be scheduled by eNB i is high, then the instantaneous estimated throughput $\Gamma_{ij}(k)$ gets the higher priority. By jointly considering the value of $\overline{R}_{ij}(k)$, $\Gamma_{ij}(k)$ and $P_{col}(i)$, the $V_{ij}[k]$ measurement captures the effect of MAC access mechanism used by eNB i which lead to precise throughput estimation.

Cluster formation

Due to varying channel conditions in every TTI, the throughput perceived from eNB i over different TTI intervals may exhibit spiky behavior. To capture this effect, we have adopted the HAC [81]. The HAC is a clustering approach for analyzing data having hierarchal relationship among them. To capture the spiky behavior of user perceived data rate in UDN scenario, here we develop a 3-step HAC algorithm. The proposed HAC algorithm classifies the data rate samples into three clusters containing *low*, *average* and *high* data rate samples.

Using HAC, in the first step we construct a *component-attribute* data matrix from $V_{ij}[\,]$. We consider the TTI period k as component and the throughput estimated at k-th TTI period $V_{ij}[k]$ as attribute. Each such component-attribute pair $(k, V_{ij}[k])$ constitutes a *node*.

Then, based on these nodes, we compute the *resemblance matrix* consisting of resemblance coefficients for every pair of nodes. Resemblance coefficient for a given pair of nodes indicate the degree of similarity or dissimilarity between those nodes, i.e., the extent the considered data rate samples are close to each other. We compute the resemblance matrix A as follows.

$$A = [a_{x,y}]_{l \times l} \tag{3.5}$$

where $a_{x,y}$ is the resemblance coefficient between nodes $(x, V_{ij}[x])$ and $(y, V_{ij}[y])$ where $1 \leq x \leq l$ and $1 \leq y \leq l$. We compute $a_{x,y}$ as $a_{x,y} = abs(V_{ij}[x] - V_{ij}[y])$, where $abs(.)$ represents the absolute value function. A smaller $a_{x,y}$ value indicates that the nodes $(x, V_{ij}[x])$ and $(y, V_{ij}[y])$ are highly similar and vice-versa.

In the next step, we determine the clusters based on the resemblance matrix. Initially, each cluster consists of a single node. In each iteration, the *minimal* coefficient in the resemblance matrix is identified, corresponding clusters are merged and the resemblance matrix is updated by the HAC algorithm. Here, we use single linkage (SLINK) to measure the similarity between two clusters. In SLINK, the similarity measure between two clusters is defined as the minimum resemblance coefficient among all pair of nodes belonging to the considered pair of clusters. This process is continued until we have three clusters namely θ_1, θ_2 and θ_3 which corresponds to low, average and high throughput values. Based on these clusters, in the next phase, we compute the estimated throughput value from eNB i.

Throughput computation

In this step, we first compute the weights for the elements belonging to different clusters based on cluster sizes. Then based on these weights we compute the final throughput value as the weighted sum of the throughput samples. Here weight of a throughput sample is proportional to the size of the corresponding cluster. Denoting by n_α the number of nodes in cluster θ_α $(1 \leq \alpha \leq 3)$, weights for the elements belonging to cluster θ_α have been computed as:

$$w_\alpha = \frac{n_\alpha}{\sum_{\beta=1}^{3} n_\beta} \tag{3.6}$$

Denoting by $e_{x\alpha}$ the x-th element of cluster θ_α $(1 \leq x \leq n_\alpha)$, the estimated throughput C_{ij} for MT j from eNB i has been computed as weighted sum of θ_α and $e_{x\alpha}$ as follows:

$$C_{ij} = \sum_{\alpha=1}^{3} \left(w_\alpha \sum_{x=1}^{n_\alpha} e_{x\alpha} \right) \tag{3.7}$$

Algorithm 1: Proposed throughput estimation algorithm

Input : $P_{col}(i)$, $\overline{R}_{ij}(k)$, $\Gamma_{ij}(k)$ and l.
Output: C_{ij}, the estimated throughput value for MT j from eNB i.

/* Sample vector preparation */
for $k = 0; k \leq l; k = k + 1$ **do**
  $V_{ij}[k] = P_{col}(i)\overline{R}_{ij}(k) + (1 - P_{col}(i))\,\Gamma_{ij}(k)$
end
/* Cluster formation based on HAC */
Determine the component-attribute data matrix $V_{ij}[\,]$.
Determine the resemblance matrix A based on $V_{ij}[\,]$.
Determine clusters θ_1, θ_2 and θ_3 based on A using SLINK.
Determine the sizes n_1, n_2 and n_3 of θ_1, θ_2 and θ_3 respectively.
/* Throughput computation */
for $\alpha = 0; \alpha \leq 3; \alpha = \alpha + 1$ **do**
  $w_\alpha = \frac{n_\alpha}{\sum_{\beta=1}^{3} n_\beta}$
end
$C_{ij} = \sum_{\alpha=1}^{3} \left(w_\alpha \sum_{x=1}^{n_\alpha} e_{x\alpha} \right)$

The proposed throughput estimation algorithm in psudocode format is presented in algorithm 1.

Time complexity

Since we are employing HAC algorithm for throughput estimation, time required to estimate the achievable throughput by MT j from eNB i is $O(l^3)$. Now, let us consider that the serving network is a *unit* circle and user density in the considered scenario is Λ_u. Hence, the number of MT requesting handover to the serving cell is $O(\Lambda_u)$. As per the definition of UDN [82], the network density should be comparable to the user density to exploit the potential of dense deployment scenario. Accordingly, the number of neighboring access networks of each MT is also $O(\Lambda_u)$. Hence, the total time required by the serving cell to estimate throughput from neighboring networks for all the MTs is $O(\Lambda_u^2 l^3)$. Based on our proposed throughput estimation algorithm, in subsequent sub-

sections, we propose a handover decision metric namely service goodness (SG) and an efficient blackout discovery mechanism.

3.3 Notion of Service Goodness (SG)

In this section, we analytically present our proposed decision metric SG for target network selection in UDN scenario. The notion of SG of an access network represents the *possibility* of getting the requested rate from the concerned access network by an MT. The proposed metric makes use of our proposed throughput estimation mechanism in order to capture the effect of MAC scheduling details.

Service goodness G_{ij} of eNB i with respect to MT j is defined as a *state variable* which represents the *possibility* of getting the requested rate r_j^{req} from eNB i. Here $G_{ij} \in \{ safe, unsafe, descend, intermittent \}$. The value of G_{ij} is determined based on C_{ij}, the estimation of achievable throughput by MT j from eNB i and $C_{ij}^{'}$, the rate of change of user perceived throughput at MT j from eNB i. Here, we compute $C_{ij}^{'}$ from $V_{ij}[\,]$ using *Newton's backward interpolation* formulae for numerical differentiation as follows:

$$C_{ij}^{'} = \nabla y_l + \frac{1}{2}\nabla^2 y_l + \frac{1}{3}\nabla^3 y_l + \frac{1}{4}\nabla^4 y_l \ldots \tag{3.8}$$

where ∇ is the standard *backward difference operator*. Here ∇y_k is computed as: $\nabla y_k = V_{ij}[k] - V_{ij}[k-1]$, where $1 \leq k \leq l$. Positive value of $C_{ij}^{'}$ indicates that eNB i is highly probable to provide r_j^{req}, on the other hand a negative value of $C_{ij}^{'}$ indicates eNB i is less probable to provide r_j^{req}. Based on $C_{ij}^{'}$ and C_{ij} the current estimated throughput value from eNB i by MT j, G_{ij} has been defined as follows. The value of G_{ij} is *safe* if $C_{ij} \geq r_j^{req}$ and $C_{ij}^{'} > 0$, i.e., eNB i is highly probable to provide r_j^{req} in near future. In contrast, G_{ij} is *unsafe* if $C_{ij} \leq r_j^{req}$ and $C_{ij}^{'} \leq 0$, i.e., eNB i is less probable to provide r_j^{req}. Here G_{ij} becomes *descend* if $C_{ij} > r_j^{req}$, but may not be able to provide the rate in near future, i.e., $C_{ij}^{'} \leq 0$. Finally, G_{ij} is *intermittent* if $C_{ij} \leq r_j^{req}$ but is highly probable to provide the data rate in near future. In brief, G_{ij} can be defined as follows.

$$G_{ij} = \begin{cases} safe, \text{ if } C_{ij} \geq r_j^{req} \text{ and } C'_{ij} > 0. \\ unsafe, \text{ if } C_{ij} < r_j^{req} \text{ and } C'_{ij} \leq 0. \\ descend, \text{ if } C_{ij} \geq r_j^{req} \text{ and } C'_{ij} \leq 0. \\ intermittent, \text{ if } C_{ij} < r_j^{req} \text{ and } C'_{ij} > 0. \end{cases}$$

3.4 Proposed Blackout Discovery Mechanism

The role of this phase is to predict an upcoming blackout period. The blackout discovery phase makes use of our proposed throughput estimation algorithm. In this phase, an MT periodically measures the estimated time after which a blackout period is going to start. Based on the values of C_{ij} and C'_{ij}, the estimated time E_{ij} that MT j will get r_j^{req} from eNB i can be computed as follows:

$$E_{ij} = \frac{C_{ij} - r_j^{req}}{abs(C'_{ij})} \tag{3.9}$$

where $abs(.)$ represents the absolute value function. The blackout discovery phase triggers handover if:

$$E_{ij} \leq d^{mac} \tag{3.10}$$

where d^{mac} is the average MAC layer delay required to request and acquire a channel from the target access network.

Based on the proposed blackout discovery and notion of SG, in the next section we propose a SG based handover algorithm (SGHO) for UDN scenario.

3.5 Proposed Service Goodness Based Handover Algorithm (SGHO)

The proposed SGHO algorithm consists of three phases. In *candidate network discovery* phase, MT j periodically determines the set of candidate networks based on RSRP mea-

surements. The *handover decision* phase starts as soon as an upcoming blackout period is predicted by the proposed blackout discovery mechanism. In this phase, the target network is determined from the set of candidate networks based on their SG values. Finally, in *handover completion* phase, all active sessions are switched from current network to the target network.

3.5.1 Candidate network discovery

In this phase, MT j constructs the set of candidate networks $O(j)$ based on RSRP measurements. Assuming that $O(j)$ includes m' number of eNBs ($m' < m$), we represent $O(j)$ as:

$$O(j) = [\texttt{eNB } 1, \texttt{eNB } 2, \ldots, \texttt{eNB } m']$$

It may be noted that eNBs belonging to $O(j)$ can be both MeNB and SeNBs.

3.5.2 Handover decision

This phase starts as soon as an upcoming blackout period is predicted and handover is triggered by our proposed blackout discovery mechanism. The handover decision phase is presented in flowchart form in Fig. 3.2. In this phase, first the SG values of the candidate networks belonging to $O(j)$ are computed and then the target network is determined based on the SG values. We represent $\Omega(j)$ the set of SG values of the candidate networks belonging to $O(j)$ as follows:

$$\Omega(j) = [G_{1j}, G_{2j}, \ldots, G_{m'j}]$$

where G_{ij} is the SG value of eNB i ($1 \leq i \leq m'$) with respect to MT j.

To determine the best network from $\Omega(j)$, we partition $\Omega(j)$ based on the SG values of the candidate networks. The partition consists of four subsets namely S, U, D and I which contain the access networks having the SG values *safe*, *unsafe*, *decay* and *intermittent* respectively. It may be noted that these subsets are mutually exclusive and exhaustive. After computing the partition, target network selection is performed as follows. If S is non-empty, then an access network from S is selected *randomly*. If S is empty but I is non-empty, then we defer the decision for $q \times T_s$ time period where q is a

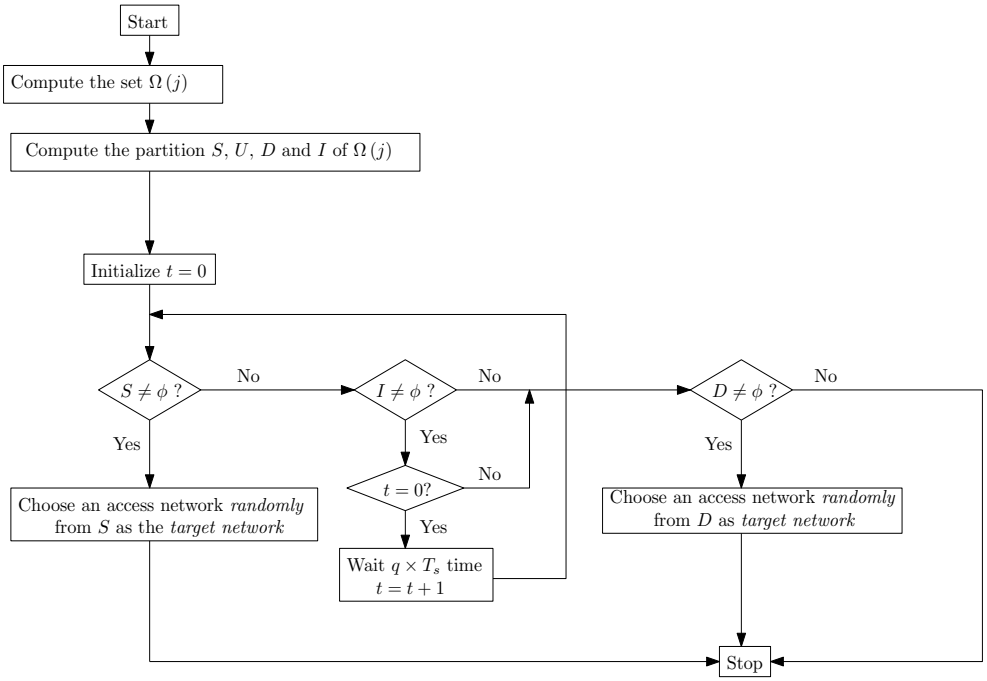


Figure 3.2: Handover decision phase of SGHO algorithm

positive integer and T_s is the sampling interval of RSRP measurements used in candidate network discovery phase. This is because, it is highly probable that an access network of I will have a SG value *safe* after $q \times T_s$ time period, i.e., the set S may become non-empty. To keep track of the fact that whether MT j has waited $q \times T_s$ time period when $I \neq \phi$, we have used a binary variable namely Flag. If S is found non-empty after $q \times T_s$ time period, then an access network is selected from S *randomly* as the target network. Otherwise, we look for the set D. If D is non-empty, then an access network is selected randomly from D, else we randomly select an access network from I as the target network. If all of the sets S, D and I are empty then the call is dropped as the remaining candidate networks are unsafe, i.e., those networks are unable to provide the rate at present and in near future as well.

3.5.3 Handover completion

In this phase, resources are released from the current network and all active sessions are switched from current network to the target network.

In a previous study [83], it has been shown that a single access network can provide the requested data rate as long as the traffic load in the system is below a particular limit. When the traffic load increases beyond that limit, then a single access network can not provide the requested data rate due to scarcity of RBs. In that situation, it is worthy to combine the data rates perceived from multiple access networks in order to satisfy the requested data rate. To deal with the blackout period under high traffic load condition, in the next section, we propose an SHT which can be used with any of the existing handover algorithms.

3.6 Proposed Soft Handover Technique

The SHT is invoked as soon as an upcoming blackout period is encountered by our proposed blackout discovery mechanism. Our proposed SHT can be used with any existing handover algorithm to minimize the blackout period. The proposed SHT consists of two phases. In *throughput enhancement* phase, data rates perceived from multiple eNBs are combined using DC support at the user end. In *access network switch* phase, all the active sessions of MT j are switched to the target eNB as soon as the estimated throughput from the target eNB become higher than r_j^{req}.

3.6.1 Throughput enhancement

The throughput enhancement phase utilizes DC support as well as our proposed throughput estimation mechanism. In this phase, data rates perceived from serving as well as target eNBs are aggregated at the receiver end using DC mechanism. To illustrate the throughput enhancement phase, let us consider that MT j is currently associated with eNB i and is getting the data rate δ_{ij} which is less than r_j^{req}. Let eNB t has been selected as the target eNB depending on a predefined decision metric. In order to get r_j^{req} in DC mode, MT j needs to get atleast $r_j^{req} - \delta_{ij}$ from eNB t. Here MT j switches to the DC mode if C_{tj}, the estimated throughput from eNB t by MT j is greater than $r_j^{req} - \delta_{ij}$. In this phase, a binary variable Temp is used to indicate whether MT j can get r_j^{req} from eNB i and eNB t in DC mode. The call is dropped if MT j can not get r_j^{req} in DC mode, i.e., $C_{tj} < r_j^{req} - \delta_{ij}$.

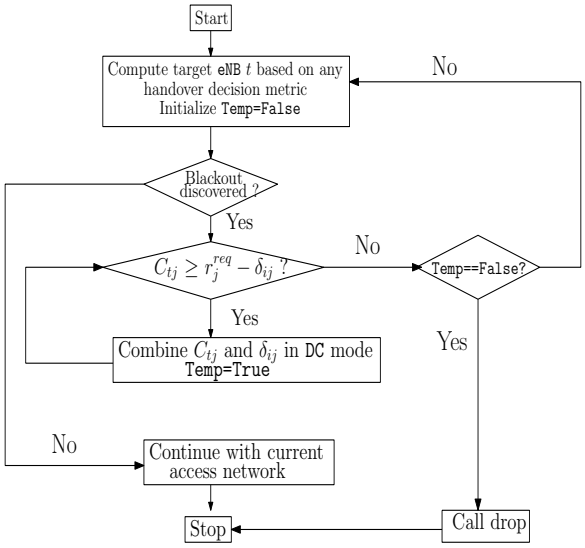


Figure 3.3: Flowchart of the proposed SHT algorithm

This phase adopts the 3C architecture of DC where the serving eNB acts as the *master-cell* and the neighboring target eNB as *secondary-cell*. Here, a single packet data convergence protocol (PDCP) layer is located at the master-cell, however radio link control (RLC) layers are present in both the master-cell and the secondary-cell. In this architecture, an MT is served by a single bearer and its flow split occurs at the master-cell. In DC mode, an MT can utilize resources across both master-cell and secondary-cell for the same bearer. Such resource utilization increases the per-user throughput for a given application.

3.6.2 Access network switch

All active sessions are switched to eNB t as soon as the time average value of throughput perceived from eNB t become sufficient to meet r_j^{req}, i.e., $C_{tj} \geq r_j^{req}$. The overall operation of the proposed SHT algorithm in flowchart format is shown in Figure 3.3.

In the next section, we compare the performances of our proposed SHT with that of semisoft handover mechanism based on Markov model analysis. We consider the steady state probabilities of experiencing blackout period as performance evaluation metrics.

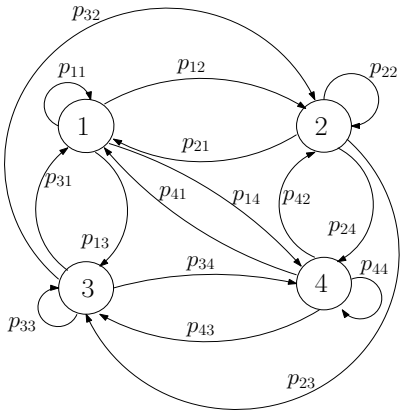


Figure 3.4: Markov model for the proposed SHT

3.7 Performance Analysis of Proposed SHT and Semisoft handover

In this section, we develop a Markovian model to trace the system behavior when SHT and semisoft handover are used. It may be noted that the actual system behavior is highly stochastic as it involves several random factors such as fading, user mobility, call arrival rate and interference in overlay networks. For the sake of tractability, we approximate the system behavior through a Markov process. The states of the proposed Markov based model is mainly concerned about whether the MT is getting the requested rate only from MeNB/SeNB or both. The past effect of all other factors (e.g., user mobility) is encoded in the present state of the Markov process. The future state in turn depends only on the present state and thereby satisfy the Markov property. In subsequent subsections, we analyze the probability of experiencing blackout by an MT based on Markov models for both the proposed SHT and semisoft handover mechanisms.

3.7.1 Markov model for the proposed SHT

Our proposed SHT can provide the requested data rate by combining data rates from MeNB and SeNB at the cell edge when neither MeNB nor the SeNB alone can provide the requested data rate. While roaming across the considered UDN scenario, an MT can reside in one of the four states:

- State 1: the state that the MT is getting the requested data rate entirely from MeNB.
- State 2: the state that the MT is getting the requested data rate entirely from SeNB.
- State 3: the state that neither serving MeNB nor target SeNB alone can provide the requested data rate to the MT, however the data rate service can be provided in DC mode by combining the data received from current and target eNBs.
- State 4: the state that neither serving MeNB nor target SeNB is providing the requested data rate to the MT, and the requested data rate can not even be provided in DC mode.

Denoting by $p_{i'j'}$ the transition probability from state i' ($1 \leq i' \leq 4$) to state j' ($1 \leq j' \leq 4$), the transition probability matrix Z can be defined as follows.

$$Z = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \tag{3.11}$$

The resulting Markov model has been shown in Figure 3.4. It may be noted that for all pair of possible states (i', j') in the considered Markov model, if state j' is reachable from state i', then state i' is also reachable from state j'. Hence the Markov chain has a *single recurrent class*. Markov chain is *aperiodic* because there exists no integer f such that the set of all states can be partitioned into disjoint subsets $W_1, W_2, \ldots, W_f$ in such a way that for all $p_{i'j'} > 0$ the following conditions hold: (1) If $i' \in W_{k'}$ then $j' \in W_{k'+1}$, $1 \leq k' \leq f-1$ and (2) if $i' \in W_f$ then $j' \in W_1$. Since the Markov chain is aperiodic and has only one recurrent class, from *Chapman-Kolmogorov* equations we get:

$$\pi_{j'} = \sum_{k=1}^{4} \pi_k p_{kj'} \quad \text{and} \quad \sum_{j'=1}^{4} \pi_{j'} = 1 \tag{3.12}$$

where $\pi_{j'}$ is the steady state probability of reaching state j' starting from any state.

Clearly, π_4 represents the steady state probability that an MT will experience a black-out period. Here π_4 can be computed by solving the system of linear equations (3.12). The final expression for π_4 has been depicted in Equation (3.13).

3.7.2 Markov model for semisoft handover

On the other hand, in semisoft handover mechanism, an MT can get the requested data rate either from MeNB or from SeNB [15]. The Markov model for semisoft handover consists of three states. State 1 and state 2 are common for both SHT and semisoft handover. It may be noted that unlike SHT, an MT can not receive data from multiple access network simultaneously in order to satisfy the data rate request. Hence, state 3 and state 4 of SHT is merged in one state, i.e., state 3 in semisoft handover. Denoting by $p'_{i'j'}$ the transition probability from state i' to state j' for semisoft handover mechanism, the transition probability matrix Z' can be written as:

$$Z' = \begin{bmatrix} p'_{11} & p'_{12} & p'_{13} \\ p'_{21} & p'_{22} & p'_{23} \\ p'_{31} & p'_{32} & p'_{33} \end{bmatrix} \tag{3.14}$$

Since state 1 and state 2 are identical in both the models for SHT and semisoft handover mechanism, $p'_{11} = p_{11}$, $p'_{12} = p_{12}$, $p'_{21} = p_{21}$ and $p'_{22} = p_{22}$ holds. On the other hand, state 3 and state 4 of SHT is merged in one state, i.e., state 3 in semisoft handover. Hence $p'_{13} > p_{13}$ and $p'_{23} > p_{23}$ holds good. Consequently, transition probabilities to state 3 increases. The resulting Markov chain is shown in Figure 3.5. Similar to the Markov chain for SHT, the Markov chain for semisoft handover is aperiodic and also has one recurrent class. Hence, applying Chapman-Kolmogorov equation we get:

$$\pi'_{j'} = \sum_{k=1}^{3} \pi'_k p_{kj'} \quad \text{and} \quad \sum_{j'=1}^{3} \pi'_{j'} = 1 \tag{3.15}$$

Here $\pi'_{j'}$ denotes the steady state probability of reaching state j' starting from any state.

$$\pi_4 = 1 - \frac{p_{14}}{a_{11}} - \frac{M_1}{K_1}\left(1 - \frac{a_{12}}{a_{11}}\right) - \frac{L_1}{K_1}\left(\frac{M_2L_1 - M_1K_2}{L_2K_1 - K_2L_1}\right) - \left(\frac{M_2K_1 - M_1K_2}{L_2K_1 - K_2L_1}\right)\left(1 - \frac{a_{23}}{a_{11}} - \frac{L_1}{K_1}\right) \tag{3.13}$$

where $a_{11} = 1 - p_{11} + p_{14}$, $a_{13} = p_{14} - p_{13}$, $a_{21} = p_{24} - p_{21}$, $a_{12} = p_{14} - p_{12}$, $a_{22} = 1 - p_{22} + p_{24}$, $a_{31} = p_{34} - p_{31}$, $a_{23} = p_{24} - p_{23}$, $a_{32} = p_{34} - p_{32}$, $a_{33} = 1 - p_{33} + p_{34}$, $M_1 = \frac{p_{24} - p_{11}p_{24} + p_{14}p_{21}}{1 - p_{11} + p_{14}}$, $K_1 = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11}}$, $K_2 = \frac{a_{11}a_{32} - a_{12}a_{31}}{a_{11}}$, $L_1 = \frac{a_{11}a_{23} - a_{21}a_{13}}{a_{11}}$, $L_2 = \frac{a_{11}a_{33} - a_{31}a_{13}}{a_{11}}$ and $L_2 = \frac{p_{34} - p_{11}p_{34} + p_{31}p_{14}}{1 - p_{11} + p_{14}}$.

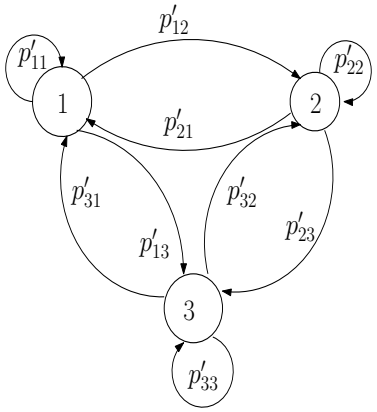


Figure 3.5: Markov model for semisoft handover

Clearly, π'_3 represents the steady state probability that an MT will experience a blackout period. Here also π'_3 can be computed solving the system of linear equations (3.15). The final expression for π'_3 is depicted in Equation (3.16).

3.8 Results and Discussions

3.8.1 Simulation setup

To evaluate the performance of our proposed SGHO algorithm as well as the proposed SHT, we have prepared a MATLAB based simulator to carry out system level simulations. We consider a LTE-A HetNet scenario where 7 MeNBs are providing ubiquitous coverage in a 2 km $\times$ 2 km area. As per the specification given in 3rd generation partnership project (3GPP) release 12 [5], the minimum distance between MeNBs has been set to 500 meters. Within the coverage region of the MeNBs, 28 SeNBs are randomly deployed in condensed clusters. Each cluster consists of 4 SeNBs and has a circular area with 50 meters radius. Here minimum distance between SeNBs has been set to 20 meters.

$$\pi_3^{'} = 1 - \left[\frac{p'_{14}}{1 - p'_{11} + p'_{14}} + \frac{M'_1}{K'_1}\left(1 - \frac{a'_{12}}{a'_{11}}\right)\right] \tag{3.16}$$

where $a'_{11} = 1 - p'_{11} + p'_{13}$, $a'_{12} = p'_{13} - p'_{12}$, $a'_{21} = p'_{23} - p'_{21}$, $a'_{22} = 1 - p'_{22} + p'_{23}$, $M'_1 = \frac{p'_{23} - p'_{11}p'_{23} + p'_{13}p'_{21}}{1 - p'_{11} + p'_{13}}$ and $K'_1 = \frac{a'_{11}a'_{22} - a'_{12}a'_{21}}{a'_{11}}$.

Transmitting powers of the MeNB and SeNBs have been set to 46 dBm and 23 dBm respectively [84]. To compute the path loss in MeNBs, we have employed the propagation model for urban and suburban areas. Here pathloss in dB is computed as $Pathloss = 128.1 + 37.6 \log_{10} R$ where R is distance between the MeNB and the corresponding MT in kilometers [85]. In SeNBs, we compute the path loss in dB as $Pathloss = 24 + 45 \log_{10}(R + 20)$ [85]. Total bandwidth for the MeNB and the SeNBs have been set to 10 MHz and 5 MHz respectively. Both MeNB and SeNBs are assumed to use PF as MAC access mechanism. In our considered simulation scenario, MTs are assumed to be *uniformly distributed* and roaming according to *smooth random waypoint* mobility model [79]. The velocity of an MT can vary from 3 km/h (pedestrian) to 100 km/h (high mobile). Acceleration of the MTs ranges from $0\ m/s^2$ to $5\ m/s^2$. An MT accelerates or decelerates depending on its current and target velocity. MTs are assumed to have data rate requests for constant bit rate (CBR) *youtube* video traffic and variable bit rate (VBR) *videoconferencing* traffic. Here CBR and VBR traffic have been modeled similar to that of [51]. CBR traffic are assumed to arrive following a *Poisson distribution* with arrival rate λ. The average holding time of CBR traffic is *exponentially distributed* and its mean is normalized to *unity*. On the other hand, VBR traffic is assumed to arrive based on a *Pareto distribution* with parameters α_{on}, α_{off} and α_s. Here α_{on} and α_{off} are shaping parameters whereas α_s is the scale parameter. Considered parameter values are depicted in Table 3.2.

3.8.2 Performance evaluation metrics

To compare the performances of our proposed SGHO with traditional RSRP based and reference base station efficiency (RBSE) based [35] approaches, we have considered *user throughput* and *system throughput* as performance evaluation metrics. User throughput

Table 3.2: Parameter settings to evaluate the proposed SGHO and SHT

Parameter	Value	Parameter	Value
λ [51]	$6 \sim 16$	S_r [59]	12
S_e [59]	10 symbols	T_s [59]	1 ms
α_{on} [51]	1.1	α_{off} [51]	1.5
α_s [51]	1.0		

is defined as the amount of data communicated per unit time through downlink data channels while roaming across the considered LTE-A HetNet. System throughput is defined as the absolute amount of data communicated to the MTs per unit time by the system. It may be noted that these two metrics are not sufficient to measure *how long an MT is getting its requested data rate* as those metrics are only concerned about the absolute amount of data communicated per unit time. To measure the quality of service experienced by an MT, we introduce an additional metric namely *user goodness*. We define *user goodness* as the fraction of time that an MT is getting its requested data rate during simulation time interval. This metric is particularly important to measure the quality of service for video applications. To compare the performances of our proposed SHT with that of semisoft handover, we have considered throughput and *call dropping probability* as performance evaluation metrics. Call dropping probability is defined as the probability of a call being dropped due to throughput degradation at the cell edge. Here, we assume that a call is being dropped by an MT if it is not getting the requested rate for a predefined period of time. This time interval depends on the concerned data rate service. For video traffic application, the time interval is typically set to 300 ms [86].

3.8.3 Performance evaluation of the proposed throughput estimation mechanism

Figure 3.6 depicts the effect of traffic load on blackout period based on simulation results. Here traditional hard handover has been used for radio link management at the cell edges. The result shows that the blackout period monotonically increases with increasing traffic load. This is because the distance upto which an eNB can provide the requested data rate decreases with increasing traffic load due to the scarcity of RBs as investigated in [83].

Figure 3.7 shows that throughput estimated by our proposed algorithm is much closer to the actual throughput compared to that of SINR based estimation. This is because, unlike SINR based estimate, our algorithm explicitly considers the effect of MAC scheduling mechanism and rapid fluctuation of user perceived throughput. Such spiky behavior of maximum achievable throughput in UDN scenario has been depicted in Figure 3.8. Here we have traced the maximum achievable throughput of a user for a period of 20

TTIs while keeping its position fixed. It may be observed that the achievable throughput fluctuates over a wide range of values (794 Kbps to 3244 Kbps). This is because the interference received by an MT in OFDMA networks primarily depends on the overlapping sub-carrier allocation in nearby access networks. Such sub-carrier allocations in turn are critically influenced by the variable traffic load in UDN scenario.

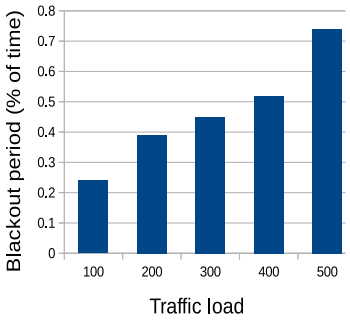


Figure 3.6: Blackout period vs. traffic load

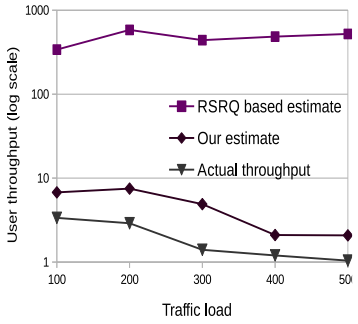


Figure 3.7: Evaluating the proposed throughput estimation mechanism

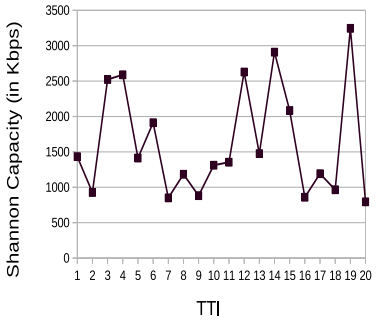


Figure 3.8: Spiky behavior of user throughput

3.8.4 Performance evaluation of SGHO

In this section, we compare our proposed SG based SGHO algorithm with that of traditional RSRP based [87] and reference base station efficiency (RBSE) based [35] handover algorithms. The traditional RSRP based handover is also known as 3GPP A3 handover. This algorithm collects RSRP measurements from the serving as well as the neighboring cells in every TTI. The handover process is triggered when RSRP of a neighboring cell becomes higher than the serving cell, i.e., A3 event occurs. On the other hand, the RBSE metric accounts the cumulative effect of transmission power, traffic load and the users' spectral efficiency to select the target network. Here also the handover initiation is based on A3 event. In this study, we consider the data rate requests for CBR *youtube* video standards. Here the considered data rates are 1.5 Mbps (for YouFlash), 0.2 Mbps (for YouHtml), 2.5 Mbps (for YouHD) and 2.7 Mbps (for YouMob) [88]. An MT can have data rate request for any one of these video standards with equal probability.

Figures 3.9-3.11 depict the effect of traffic load on various performance evaluation metrics. Total number of MTs in the system varies from 100 MTs to 500 MTs with a step of 100 MTs. Figure 3.9 shows that the proposed SGHO algorithm significantly outperforms the RSRP based and RBSE based handover algorithms in terms of user throughput. The underlying impetus of this phenomenon is efficient blackout discovery mechanism of SGHO, and better QoS-awareness of the proposed SG compared to that of RSRP and RBSE. It is to be noted that the user perceived throughput depends on the interference level of the received signal, user density and the MAC scheduling mechanism of the serving access network. In RSRP based approaches, usually the MeNB is selected as the target network due to its higher transmission power. As a result, the MeNB become highly loaded whereas the SeNBs remain underutilized. Increasing traffic load in LTE-A HetNet cause higher sub-carrier collision probability which in turn increases the interference level of the received signal. As a result, target network selection based on RSRP causes serious degradation of user perceived throughput. In contrast to the RSRP based approach, the RBSE based algorithm explicitly considers the cumulative effect of transmission power, network traffic load as well as interference level of the received signal while selecting the target network. Due to this advanced QoS-awareness, the user perceived throughput in RBSE based approach is higher compared to that of RSRP based approach. Although RBSE based handover mechanism increases the user perceived throughput, it can not guarantee whether an user will get its requested data rate from the target network. This is because, RBSE do not consider the effect of MAC access mechanism and the spiky behavior of user throughput in UDN scenario. In contrast to the proposed RBSE based approach, the proposed SG explicitly considers the effect of MAC scheduling details by jointly considering the past average throughput, sub-carrier availability and channel condition. The proposed SGHO algorithm also employs a throughput estimation mechanism which captures the spiky behavior of user throughput using HAC algorithm. In addition, SGHO triggers handover as soon as an upcoming blackout period is encountered. As a result, MTs do not suffer from throughput degradation at the cell edges. Due to all such features, the proposed SGHO algorithm significantly outperforms the RSRP based (performance gain 58%-76% approximately) and RBSE based (performance gain 51%-73% approximately) approaches in terms of user perceived throughput.

The effect of traffic load has also been investigated on user goodness as depicted in

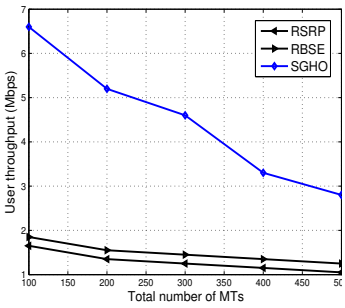


Figure 3.9: User throughput vs. traffic load

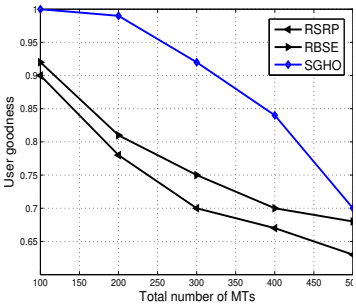


Figure 3.10: User goodness vs. traffic load

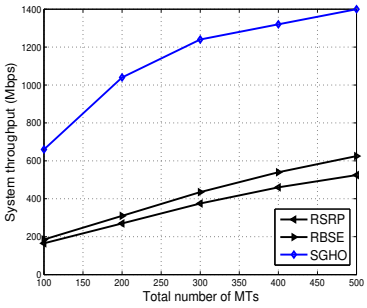


Figure 3.11: System throughput vs. traffic load

Figure 3.10. The figure shows that the proposed SGHO algorithm significantly outperforms both the RSRP based and RBSE based algorithms. The reasons behind are as follows. The achievable data rates from an access network can not be estimated precisely through RSRP or RBSE measurements. In contrast, being equipped with the efficient throughput estimation mechanism, achievable throughput from the target network can be better estimated through SG measurements. Consequently, the SGHO algorithm always selects the network which is able to provide the data rate request for a long period of time. As a result, the SGHO algorithm significantly outperforms the RSRP based (performance gain 7%-20% approximately) and RBSE based (performance gain 3%-16% approximately) approach in terms of user goodness. It may also be noted that the performance gain of SGHO over RSRP and RBSE based algorithms decreases with increasing traffic load for both user throughput and user goodness. This is because the sub-carrier collision probability increases with increasing traffic load, which in turn decreases user throughput.

Figure 3.11 depicts the effect of traffic load on system throughput. Simulation results show that the SGHO algorithm significantly improves the system throughput compared to RBSE based (1%-22% percent approximately) and RSRP based (2%-36% percent approximately) algorithms. System throughput depends on both the user throughput and the total number of MTs associated with the system, which in turn depend on load balancing among different access networks. The RBSE based algorithm outperforms the RSRP based approach because the RBSE measurements lead to better utilization of SeNBs. Although RBSE based approach provides fairer load distribution, it can not address the

strict throughput requirements of the MTs. This is because RBSE measurement do not consider availability of sub-carriers as well as MAC scheduling information. On the other hand, being equipped with MAC access mechanism and sub-carrier collision probability, the SG based SGHO algorithm have better insight regarding achievable data rate from different candidate networks. As a result, SGHO outperforms RSRP and RBSE based approaches.

3.8.5 Performance comparison between the proposed SHT and semisoft handover mechanism

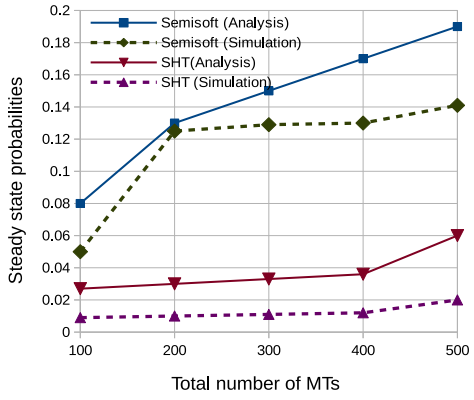


Figure 3.12: Probability of experiencing blackout vs. traffic load

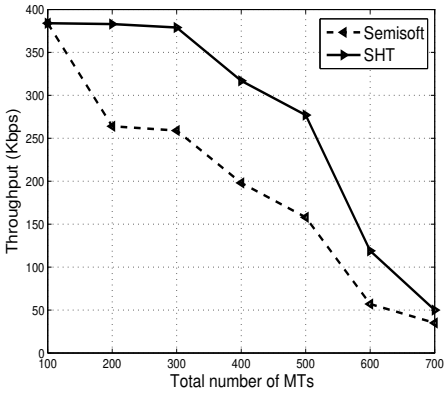


Figure 3.13: Throughput vs. traffic load (VBR traffic)

In this section, we have compared the performance of the proposed SHT with semisoft handover mechanism. In this study, network selection has been performed based on RSRP metric [35]. We first present the result obtained from the Markov models proposed in section 3.7. Then, we describe the results obtained from system level simulations where throughput and call dropping probability have been considered as performance evaluation metrics. Results obtained from analyses and simulations generate a consensus that our proposed SHT significantly outperforms the existing semisoft handover mechanism.

Figure 3.12 depicts the effect of traffic load on steady state probabilities of experiencing blackout period in semisoft handover mechanism and our proposed SHT. In

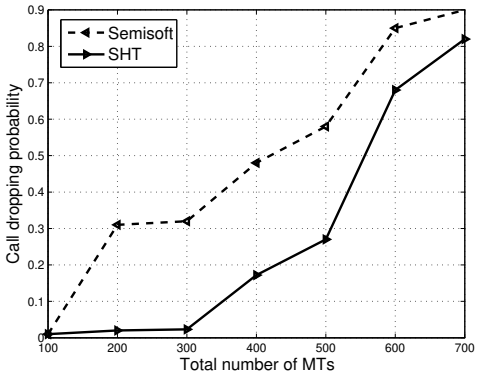


Figure 3.14: Call dropping probability vs. traffic load (VBR traffic)

this study, the transition probabilities $p_{i'j'}$ have been estimated from simulation traces as follows:

$$p_{i'j'} = \frac{\text{Number of transitions from state } i' \text{ to state } j'}{\text{Number of transitions from state } i' \text{ to any other state}}$$

Analytical results obtained from our proposed Markov model have also been validated against simulation results. It may be observed that both analytical and simulation results are following similar trends. Results show that the probability of experiencing blackout period in our proposed SHT is much lower compared to that of semisoft handover mechanism. This is because, in semisoft handover an MT experiences blackout when none of the access networks alone can provide the data rate request. On the other hand, our proposed SHT can combine data rates from current as well as target access networks when neither of the networks alone can provide the data rate request. Since, data rate from multiple access networks can not be combined in semisoft handover, the probability of experiencing blackout in SHT is much lower compared to that of semisoft handover mechanism.

Figures 3.13 and 3.14 demonstrate the performance gain of SHT over semisoft handover mechanism when the system load varies from 100 MTs to 700 MTs with a step of 100 MTs. Figure 3.13 depicts the effect of traffic load on goodput. The results show that the proposed SHT significantly outperforms the semisoft handover mechanism (performance gain 2% to 60% percent approximately) when the traffic load is in between 100 MTs to 500 MTs. In this study, we consider the data rate requests for VBR *videoconferencing* whose data rate request varies from 20 Kbps to 320 Kbps [89]. This is because,

in LTE−A HetNet, the sub-carrier collision probability increases with an increase in traffic load which in turn increases the inter-cell interference. Consequently, the perceived throughput from an access network decrease when the traffic load increases from 100 MTs to 500 MTs. An MT is not able to get the requested data rate from a single access network at cell edges during handover and frequently encounters blackout periods. This essentially invokes the proposed SHT which can predict the downfall of user perceived throughput through blackout discovery mechanism and enhances the performance in DC mode. It may be noted that the proposed SHT does not provide significant gain when the traffic load is too low (< 100 MTs) or too high (> 500 MTs). This is because under low load condition, the MTs need not to invoke the SHT as most of the time they get the requested rate from a single access network. Under high traffic load condition (> 500 MTs), the SHT can not prevent the downfall of throughput below the requested rate due to scarcity of sub-carriers and increased inter-cell interference. Consequently, the performance gain shows a decreasing trend under high load condition.

Figure 3.14 depicts the effect of traffic load on call dropping probability. We consider that a call is dropped if the concerned MT gets the requested data rate neither from a single radio access network nor in DC mode for 300 ms time duration. A call can survive as long as the serving access network can allocate enough RBs to the concerned MT and the interference level of the received signal is small enough to provide the target SINR threshold. The inter-cell interference increases with increasing traffic load in LTE−A HetNet. As a result, call dropping probability shows an increasing trend with increasing traffic load. It may be observed that the proposed SHT significantly outperforms semisoft handover mechanism (performance gain upto 90%) when traffic load is in between 100 MTs to 500 MTs. It may also be noted that the SHT do not have much effect (performance gain $1\% - 2\%$) on call dropping probability when the system load is too low, i.e., < 100 MTs. The performance gain also shows a decreasing trend when the system load is very high, i.e., > 500 MTs. The reason behind is similar to that of user throughput described previously.

Figure 3.15 and 3.16 demonstrate the performance gain on different metrics under high traffic load condition. Here we set traffic load to 1000 MTs whereas the requested rate varies from 192 Kbps to 512 Kbps with a step of 64 Kbps. Figure 3.15 compares the performances of semisoft handover mechanism with that of the proposed SHT in

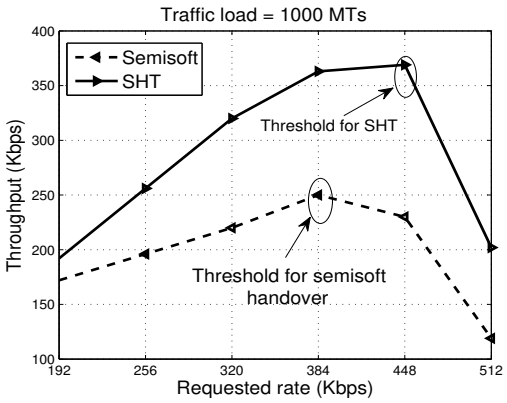


Figure 3.15: Throughput vs. requested rate (VBR traffic)

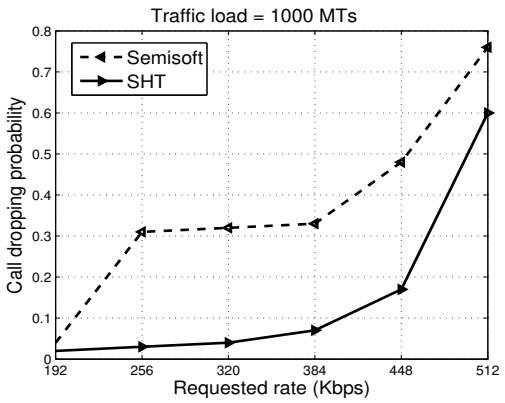


Figure 3.16: Call dropping probability vs. requested rate (VBR traffic)

terms of user perceived throughput. The result shows that the proposed SHT significantly outperforms the semisoft handover mechanism (performance gain $45\% - 50\%$) when the requested data rate is in between 192 Kbps and 448 Kbps. In LTE-A HetNet, requirements for RBs increases with increasing data rate request. As a result, duration of blackout period increases as the requested data rate increases from 192 Kbps to 448 Kbps. Consequently, the proposed SHT is invoked to deal with the blackout period. The SHT remain ineffective for lower data rate services (< 192 Kbps) as the concerned MT can get the requested data rate from a single access network. The performance gain shows a decreasing trend when the requested data rate increases beyond 448 Kbps. This performance degradation is due to the scarcity of RBs under high traffic load condition. It may also be noted that the throughput under both SHT and semisoft handover shows an increasing trend upto a threshold (384 Kbps for semisoft handover and 448 Kbps for the proposed SHT). Beyond that threshold the goodput decreases for both the cases. This is because, the system attains maximum resource utilization when the data rate request reaches that threshold. It is also important to note that the threshold for SHT is higher compared to that of semisoft handover. *This phenomenon implies that our proposed SHT ensures higher resource utilization compared to that of the semisoft handover.*

Figure 3.16 depicts the performance comparison between the proposed SHT and semisoft handover mechanism in terms of call dropping probability. The results show that the proposed SHT significantly outperforms the semisoft handover mechanism in terms of call dropping probability when the requested data rate is in between 192 Kbps

and 512 Kbps. The SHT remain ineffective for low data rate service ($<$ 192 Kbps). The performance gain decreases as the requested data rate increases beyond a certain threshold ($>$ 448 Kbps). The reason behind is similar to that described previously.

3.9 Conclusions

In this work, we have addressed the problem of blackout period in UDN scenario. We have proposed an efficient algorithm for throughput estimation explicitly considering the MAC scheduling details as well as random fluctuation of interference level in different TTIs. Based on the throughput estimation mechanism, we have proposed a decision metric SG and an efficient blackout discovery mechanism. Further, based on SG and blackout discovery mechanism, we propose a handover algorithm namely SGHO for target network selection in UDN scenario. To deal with the blackout period under high traffic load situation, we have proposed an SHT and analyzed its performance based on Markov models. From analyses and simulation results, we conclude that the proposed SGHO significantly outperforms the RSRP and RBSE based approaches in terms of user throughput and system throughput. Moreover, it has been shown that the proposed SHT can significantly reduce the blackout period compared to that of the semisoft handover.

Chapter 4

A Predictive Handover Mechanism for 5G Ultra Dense Networks

The 5G UDN is envisioned as a very dense deployment of low power eNBs where heterogeneous RATs are used to satisfy the data rate demand of MTs employing both *licensed* and *unlicensed* spectrum. In UDN scenario, conditions of the channels operating in licensed band may exhibit intermittent characteristics due to the varying level of interference received from large number of nearby access networks. On the other hand, channel conditions in unlicensed band may fluctuate drastically due to the interference caused by the co-existence of long term evolution in unlicensed band (LTE−U) cellular network and WLAN. The traditional handover mechanisms rely on the instantaneous assessments of link qualities such as RSS and SINR. The target network selected based on such instantaneous values may not be the appropriate one when the actual handover is executed. This causes in higher packet losses and severe throughput degradation. We refer to this problem as handover anomaly problem. In this chapter, we have proposed a *predictive* handover mechanism which can estimate achievable throughput values from different candidate access networks prior to handover execution. Simulation results confirm that our proposed predictive handover mechanism significantly outperforms an existing reference base station efficiency (RBSE) [35] based approach.

The rest of the chapter is organized as follows. In section 4.1, the considered network architecture has been described. In section 4.2, we present our proposed predictive handover mechanism. In section 4.3, we present the simulation results. Finally, section 4.4

Table 4.1: Important notations used to illustrate the proposed predictive handover

Notation	Meaning
$\gamma_{ij}^{l}(t)$	SINR received from LBS i by MT j at time t.
$\tau_{ij}^{l}(t)$	The traffic channel power received from LBS i by MT j at time t.
$I_{kj}^{l}(t)$	The co-channel interference received from LBS k by MT j at time t.
$N^{l}(i,j)$	The set of nearby eNBs of eNB i from which MT j is experiencing co-channel interference through licensed spectrum.
$P_{col}^{l}(i)$	Subcarrier collision probability at LBS i in licensed spectrum.
$\tau_{ij}^{u}(t)$	Traffic channel power received from UBS i by MT j at time t.
$I_{kj}^{u}(t)$	Total interference received from UBS k by MT j at time t.
$N^{u}(k,j)$	The set of all nearby eNBs of UBS k from which MT j experiences co-channel interference.
$Q(i,j)$	The set of all nearby APs of AP i from which MT j experiences co-channel interference.
$I_{kj}^{a}(t)$	The interference received from AP k by MT j at time t.
$P_{col}^{u}(i)$	The subcarrier collision probability at UBS i in unlicensed spectrum.
$\pi_{ij}^{(.)}(t)$	The achievable bit rate from eNB i by MT j at time t, $(.)$ indicates either licensed or unlicensed band.
$\psi_{ij}(t)$	Achievable bit rate by MT j from AP i at time t.
$\xi_{i}(t)$	The total number of MTs associated with AP i at time t.
$\phi_{ij}(t)$	The physical bit rate at which MT j is associated with AP i at time t.
τ	The threshold value defining the handover decision criteria.
$\delta_{j}(t)$	The time to handover for MT j at time t.
$\mu_{j}(t)$	The mean throughput vector for MT j after handover execution at time t.
$T_{j}^{n}(t)$	The normalized instantaneous throughput vector after handover at time t.
$\mu_{j}^{n}(t)$	The normalized mean vector after handover execution at time t.
Δ	Target overflow probability.
R	Distance between macrocell and picocell.

concludes the chapter. Important notations used in this chapter have been summarized in Table 4.1.

4.1 Considered network architecture

We consider a loosely coupled integration between different radio access networks as depicted in Figure 4.1. We consider that an LTE macrocell eNB is providing ubiq-

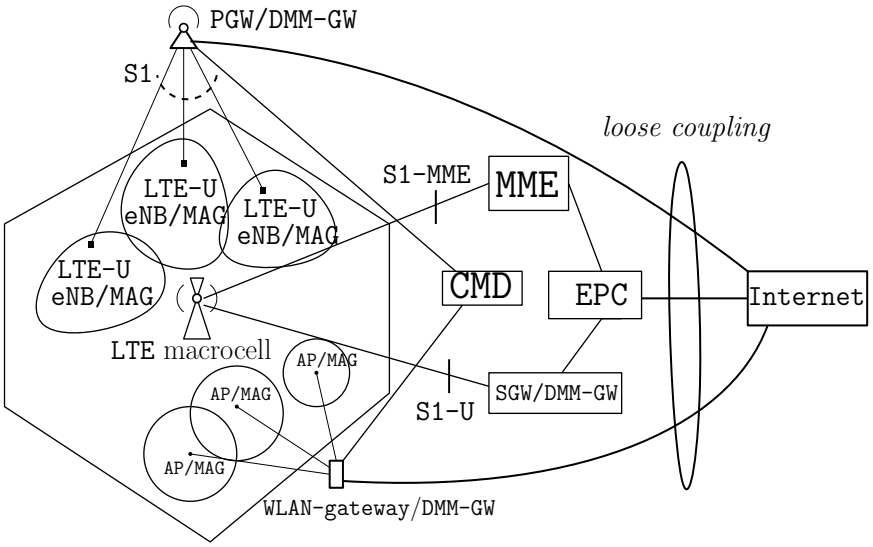


Figure 4.1: Considered network architecture

uitous coverage. Within the coverage region of the macrocell, several LTE−U picocell eNBs and IEEE 802.11n femtocell access points (APs) are deployed in an uncoordinated manner. Here the macrocell is connected to the evolved packet core (EPC) via mobility management entity (MME) and serving gateway (SGW). The macrocell is connected to MME and serving gateway (SGW) through S1−MME and S1−U interfaces respectively [90]. The picocell LTE−U eNBs are connected to a common gateway namely *picocell gateway* (PGW) through S1 interfaces [91]. Here PGW can act as concentrator, distributor as well as security gateway [91]. The IEEE 802.11n femtocell APs are also connected to a common gateway namely WLAN-gateway. The PGW, WLAN-gateway and SGW are independently connected to the Internet. Here PMIPv6 based DMM architecture [22] can be used to manage terminal mobilities. The mobile access gateways (MAGs) of PMIPv6 protocol are co-located with the LTE−U eNBs and WLAN APs. The PGW, WLAN-gateway and SGW can act as distributed mobility management gateways (DMM−GWs). The DMM−GWs are connected to the control mobility database (CMD). The CMD keeps track of current and target access networks of the MTs. In our proposed handover mechanism, the CMD assists the MTs in network selection phase.

4.2 Proposed predictive handover mechanism

Our proposed predictive handover mechanism consists of four phases namely *candidate network discovery*, *throughput estimation*, *target network selection* and *handover execution*. The network discovery and handover execution phases are based on standard methods [35, 92]. In network discovery phase, a set of candidate access networks is determined based on RSS measurements. In handover execution phase, all the active sessions are switched from the current network to the selected target network. Our main focus is on other two phases. In throughput estimation phase, throughput that the candidate networks can provide after handover execution is estimated. Based on these estimated throughput values, the target network selection problem is formulated as a stochastic integer programming (SIP) problem in the target network selection phase. The constraint set of the SIP imposes a *lower bound* on the probability that the selected target network will provide the requested data rate after handover execution. The SIP has been solved by converting the probabilistic constraint into its equivalent deterministic one using *Hoeffding bound* [93]. The functionalities of individual phases are described as follows:

4.2.1 Candidate network discovery

The candidate network discovery phase is based on background inter-frequency measurements (BIM) and relaxed measurement gap pattern [92]. We consider that, after candidate network discovery phase, the set of candidate networks for MT j contains p number of LTE−U eNBs operating in unlicensed band namely UBS 1, UBS 2, . . ., UBS p; q number of LTE−U eNBs operating in licensed band namely LBS $p+1$, LBS $p+2$, . . ., LBS $p+q$ and r number of WLAN eNBs namely WLAN 1, WLAN 2, . . . , WLAN r.

4.2.2 Throughput estimation

In OFDMA based systems, channel conditions are not significantly effected by *intra-cell* interference and the noise power is negligible compared to the power received through inter-cell interference [15]. Hence $\gamma_{ij}^{l}(t)$, the SINR received from LBS i by MT j at time

t through licensed spectrum, is given by:

$$\gamma_{ij}^{l}(t) = \frac{\tau_{ij}^{l}(t)}{P_{col}^{l}(i) \times \sum_{k \in N^{l}(i,j)} I_{kj}^{l}(t)} \tag{4.1}$$

where $\tau_{ij}^{l}(t)$ is the traffic channel power received from LBS i by MT j at time t through the allocated subcarrier in the licensed spectrum; $I_{kj}^{l}(t)$ is the co-channel interference received from LBS k by MT j at time t in licensed spectrum; $N^{l}(i,j)$ is the set of nearby LBSs of LBS i from which MT j is experiencing co-channel interference in licensed spectrum; $P_{col}^{l}(i)$ is the subcarrier collision probability which can be computed as the ratio of the total number of used subcarriers to the total number of available subcarriers operating in the licensed spectrum.

In unlicensed band communication, MT j experiences co-channel interference from neighboring LTE−U eNBs as well as WLAN APs. Following the formulation presented in [94], $\gamma_{ij}^{u}(t)$, the SINR received from AP i by MT j at time t through unlicensed spectrum, can be computed as:

$$\gamma_{ij}^{u}(t) = \frac{\tau_{ij}^{u}(t)}{P_{col}^{u}(i) \times \sum_{k \in N^{u}(i,j)} I_{kj}^{u}(t) + \sum_{k \in Q(i,j)} I_{kj}^{a}(t)} \tag{4.2}$$

where $\tau_{ij}^{u}(t)$ is the traffic channel power received from UBS i by MT j at time t through the allocated subcarrier in the unlicensed spectrum; $I_{kj}^{u}(t)$ is the total interference received from UBS k by MT j at time t in unlicensed spectrum; $N^{u}(k,j)$ is the set of all nearby UBSs of UBS k from which MT j experiences co-channel interference through unlicensed spectrum; $Q(i,j)$ is the set of all nearby APs of AP i from which MT j experiences co-channel interference; $I_{kj}^{a}(t)$ is the interference received from AP k by MT j at time t; $P_{col}^{u}(i)$ is the subcarrier collision probability in unlicensed spectrum which can be computed similar to that of $P_{col}^{l}(i)$.

Now, $\pi_{ij}^{(.)}(t)$, the achievable bit rate from eNB i (either LBS or UBS) by MT j at time t, can be computed as [95]:

$$\pi_{ij}^{(.)}(t) \;=\; \frac{b^{z} \times \gamma_{ij}^{(.)}}{H(k)} \times G_{2,3}^{3,1}\left(1, 2; k+1, 1, 1; \frac{k}{\gamma_{ij}^{(.)}}\right) \tag{4.3}$$

where $H(k) = \ln(2) \times \Gamma(k) \times k$, b^z is the bandwidth of the allocated subcarrier z, $G_{2,3}^{3,1}(.;.;.)$ is the *Meijer's G* function, $k \in [0.5, \infty)$ is a parameter representing fading severity and $(.)$ is the spectrum type indicator representing either licensed or unlicensed band. For $k = 1$, Equation (4.3) reduces to the expression of achievable throughput from *Rayleigh fading* channels. For the sake of readability, the detailed computation of $G_{2,3}^{3,1}(.;.;.)$ has been presented at the last of this chapter as an Appendix.

The achievable bit rate $\psi_{ij}(t)$ by `MT` j from `AP` i at time t in proportional fair access mechanism is given by [96]:

$$\psi_{ij}(t) = \frac{\phi_{ij}(t)}{\xi_i(t)} \tag{4.4}$$

where $\xi_i(t)$ denotes the total number of `MT`s associated with `AP` i at time t; $\phi_{ij}(t)$ denotes the physical bit rate at which `MT` j is associated with `AP` i at time t. The information regarding scheduling and load $\xi_i(t)$ can be communicated to the `MT`s by overloading the service set identifier (`SSID`) field of `802.11` beacon frame [78]. The physical bit rate $\phi_{ij}(t)$ at which `MT` j can be associated with `AP` i has been derived from the received signal strength indicator (`RSSI`) values.

Let $\theta_j(t)$ be the value of a generic metric measured from the current access network by `MT` j at time t. Note that $\theta_j(t)$ may be `RSS`, `SINR` or packet loss rate. Let τ be the corresponding threshold value defining the handover decision criteria. Hence $\delta_j(t)$, the time to handover for `MT` j at time t, is computed as:

$$\delta_j(t) = \frac{\theta_j(t) - \tau}{S_j(t)} \tag{4.5}$$

where $S_j(t)$ is the rate of change of $\theta_j(t)$ measured by `MT` j at time t. Here $S_j(t)$ has been measured using the standard techniques of moving average method [78] as follows:

$$S_j(t) = \frac{F'_j(t) - L'_j(t)}{\omega' \times T'_b} \tag{4.6}$$

where $F'_j(t)$ and $L'_j(t)$ are the average $\theta_j(t)$ measurements at `MT` j in first half and last

half of the slope estimator window ω'. Here $F'_j(t)$ and $L'_j(t)$ can be measured as:

$$F'_j(t) = \frac{2}{\omega'} \sum_{x=0}^{\frac{\omega'}{2}-1} \theta_j(t-\omega'+1+x) \text{ and}$$

$$L'_j(t) = \frac{2}{\omega'} \sum_{x=\frac{\omega'}{2}}^{\omega'} \theta_j(t-\omega'+1+x)$$

Using $\delta_j(t)$, $T_j(t')$, the instantaneous throughput vector for MT j after handover execution can be estimated as:

$$T_j(t') = \left\{\pi^u_{\alpha j}(t'), \pi^l_{\beta j}(t'), \psi_{\gamma j}(t')\right\} \tag{4.7}$$

where $t' = t + \delta_j(t)$, $\pi^u_{\alpha j}(t') = \pi^u_{\alpha j}(t) + S^u_{\alpha j}(t) \times \delta_j(t)$, $\pi^l_{\beta j}(t') = \pi^l_{\beta j}(t) + S^l_{\beta j}(t) \times \delta_j(t)$ and $\psi_{\gamma j}(t') = \psi_{\gamma j}(t) + S^a_{\gamma j}(t) \times \delta_j(t)$. Here and in the subsequent sections, $\alpha \in \{1, 2, \ldots p\}$, $\beta \in \{p+1, p+2, \ldots p+q\}$ and $\gamma \in \{1, 2, \ldots r\}$, where p, q and r are the numbers of UBSs, LBSs and number of WLAN APs respectively. Here $S^u_{\alpha j}(t)$, $S^l_{\beta j}(t)$ and $S^a_{\gamma j}(t)$ represent rate of change of $\pi^u_{\alpha j}(t)$, $\pi^l_{\beta j}(t)$ and $\psi_{\gamma j}(t)$ respectively and can be computed similar to that of $S_j(t)$. Similarly, $\mu_j(t')$, the mean throughput vector for MT j after handover execution, can be estimated similar to $T_j(t')$ as follows:

$$\mu_j(t') = \left\{\overline{\pi}^u_{\alpha j}(t'), \overline{\pi}^l_{\beta j}(t'), \overline{\psi}_{\gamma j}(t')\right\} \tag{4.8}$$

Here $\overline{\pi}^u_{\alpha j}(t')$, $\overline{\pi}^l_{\beta j}(t')$ and $\overline{\psi}_{\gamma j}(t')$ can be computed from $\overline{\pi}^u_{\alpha j}(t)$, $\overline{\pi}^l_{\beta j}(t)$ and $\overline{\psi}_{\gamma j}(t)$ respectively similar to that computed in $T_j(t')$. The values of $\overline{\pi}^u_{\alpha j}(t)$, $\overline{\pi}^l_{\beta j}(t)$ and $\overline{\psi}_{\gamma j}(t)$ represent the average values of $\pi^u_{\alpha j}(t)$, $\pi^l_{\beta j}(t)$ and $\psi_{\gamma j}(t)$ respectively over the time interval $[t-\omega', t]$. Here ω' is the slope estimator window interval of the employed moving average method. It may be noted that contextual information such as $T_j(t')$ and $\mu_j(t')$ can be periodically computed by MT j with the help of connection manager and can be communicated to the CMD through measurement reports [41, 42].

4.2.3 Target network selection

In this section, we formulate the target network selection problem as a stochastic integer program (SIP). To formulate the SIP, we introduce the following binary variables.

$$
\begin{aligned}
x_{\alpha j} &= \begin{cases} 1 \text{ if MT } j \text{ is associated with UBS } \alpha \\ 0 \text{ otherwise} \end{cases} \\
y_{\beta j} &= \begin{cases} 1 \text{ if MT } j \text{ is associated with LBS } \beta \\ 0 \text{ otherwise} \end{cases} \\
z_{\gamma j} &= \begin{cases} 1 \text{ if MT } j \text{ is associated with AP } \gamma \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
\tag{4.9}
$$

Now, the normalized instantaneous throughput vector after handover execution, $T_j^n\,(t')$, can be estimated using $T_j\,(t')$ (Equation (4.7)) as:

$$
T_j^n(t') = \left\{ \pi_{\alpha j}^{u,n}(t'), \pi_{\beta j}^{l,n}(t'), \psi_{\gamma j}^{n}(t') \right\} \tag{4.10}
$$

where $\pi_{\alpha j}^{u,n}(t')$, $\pi_{\beta j}^{l,n}(t')$ and $\psi_{\gamma j}^{n}(t')$ denote the normalized values of $\pi_{\alpha j}^{u}(t')$, $\pi_{\beta j}^{l}(t')$ and $\psi_{\gamma j}(t')$ respectively. Similarly, $\mu_j^n\,(t')$, the normalized mean vector after handover execution, can be estimated using $\mu_j\,(t')$ (Equation (4.8)) as follows:

$$
\mu_j^n(t') = \left\{ \overline{\pi}_{\alpha j}^{u,n}(t'), \overline{\pi}_{\beta j}^{l,n}(t'), \overline{\psi}_{\gamma j}^{n}(t') \right\} \tag{4.11}
$$

where $\overline{\pi}_{\alpha j}^{u,n}(t')$, $\overline{\pi}_{\beta j}^{l,n}(t')$ and $\overline{\psi}_{\gamma j}^{n}(t')$ denote the normalized values of $\overline{\pi}_{\alpha j}^{u}(t')$, $\overline{\pi}_{\beta j}^{l}(t')$ and $\overline{\psi}_{\gamma j}(t')$ respectively. Using $T_j^n(t')$ (Equation (4.10)), $\chi_j(t')$, the estimated throughput of MT j after handover execution, can be written as:

$$
\chi_j(t') = \sum_{\alpha=1}^{p} \pi_{\alpha j}^{u,n}(t') \times x_{\alpha j} + \sum_{\beta=1}^{q} \pi_{\beta j}^{l,n}(t') \times y_{\beta j} + \sum_{\gamma=1}^{r} \psi_{\gamma j}^{n}(t') \times z_{\gamma j}. \tag{4.12}
$$

The goal of this phase is to maximize $\chi_j(t')$. We consider that the radio link management

at cell edge is performed by traditional *hard* handover mechanism. Hence, the following constraint must hold:

$$\sum_{\alpha=1}^{p} x_{\alpha j} + \sum_{\beta=1}^{q} y_{\beta j} + \sum_{\gamma=1}^{r} z_{\gamma j} \leq 1. \tag{4.13}$$

It may be noted that when the left hand side of the inequality evaluates to 0, MT j communicates through the macrocell LTE eNB. Now, applying linearity of expectation and the fact that $\mathbb{E}[UV] = \mathbb{E}[U].\mathbb{E}[U]$ where U and V are independent random variables, the mean value of $\chi_j(t')$ can be computed using $\mu_j^n(t')$ (Equation (4.11)) as follows:

$$\mathbb{E}[\chi_j(t')] = \frac{1}{p}.\sum_{\alpha=1}^{p} \overline{\pi}_{\alpha j}^{u,n}(t') + \frac{1}{q}.\sum_{\beta=1}^{q} \overline{\pi}_{\beta j}^{l,n}(t') + \frac{1}{r}.\sum_{\gamma=1}^{r} \overline{\psi}_{\gamma j}^{n}(t')$$

Here each candidate network belonging to same radio access technology have been considered as equally likely to be selected as target network. To deal with the handover anomaly problem, we set a target overflow probability $\Delta \in [0, 1]$ corresponding to a desired service guarantee. This can be stated as the following probabilistic constraint:

$$\mathbb{P}\left[\chi_j(t') \geq r_j^{'}\right] \geq \Delta \Rightarrow \quad \mathbb{P}\left[\chi_j(t') \geq \mathbb{E}[\chi_j(t')] + temp(j)\right] \geq \Delta \tag{4.14}$$

where $temp(j) = r_j^{'} - \mathbb{E}[\chi_j(t')]$. Here $r_j^{'}$ is the normalized value of the requested data rate r_j^{req}. Now, to solve the SIP, we first convert the probabilistic constraint (4.14) into deterministic one using *Hoeffding's bound* [93]. The Hoeffding's bound for a series of n independent random variables Y_i $(1 \leq i \leq n)$ with known mean values $\mathbb{E}[Y_i]$ states that:

$$\mathbb{P}[Y_i \geq \mathbb{E}[Y_i] \pm \epsilon] \leq e^{\frac{-2\epsilon^2}{n}} \tag{4.15}$$

It may be noted that, the Hoeffding's bound can be applied even if the distributions of Y_is are completely *unknown*. Using Hoeffding bound, the probabilistic constraint (4.14) can be rewritten as:

$$\Delta \leq \mathbb{P}\left[\chi_j(t') \geq \mathbb{E}\left[\chi_j(t')\right] + temp(j)\right] \leq e^{\frac{-2\times temp(j)^2}{p+q+r}} \tag{4.16}$$

Now, the probabilistic constraint can be converted to its equivalent deterministic constraint as:

$$\Delta \leq e^{\frac{-2\times temp(j)^2}{p+q+r}} \Rightarrow r_j^{'} \leq \mathbb{E}\left[\chi_j(t')\right] + \sqrt{(p+q+r)\ln\left(\frac{1}{\sqrt{\Delta}}\right)} \tag{4.17}$$

So, the overall optimization problem is to maximize (4.12) subjected to the constraints (4.13) and (4.17). The solution of the formulated optimization problem is obtained by the CMD component of the considered network architecture and communicated to MT j through downlink control channels.

4.2.4 Handover execution

The handover execution phase is according to 3GPP standard [35]. Here, an MT selects a new target network from the set of available candidate networks in case of handover failure.

4.3 Results and Discussions

To evaluate the performance of the proposed predictive handover mechanism, we have prepared a MATLAB based simulator. We have compared the performances of our proposed handover mechanism with that of the RBSE based handover mechanism [35]. The RBSE metric considers the instantaneous values of transmitted power, traffic load and bandwidth requirement of the MTs to select the target network.

4.3.1 Performance evaluation metrics

To evaluate the performance of the proposed predictive handover mechanism we have considered goodput, handover failure probability and number of handover as perfor-

mance evaluation metrics. Goodput is defined as the amount of data communicated per unit time to an MT when the MT is getting the data rate request. Handover failure probability is defined as the probability that the target network will not be able to provide the requested data rate after handover execution. The number of handover indicates the normalized value of the total number of handovers during simulation time interval.

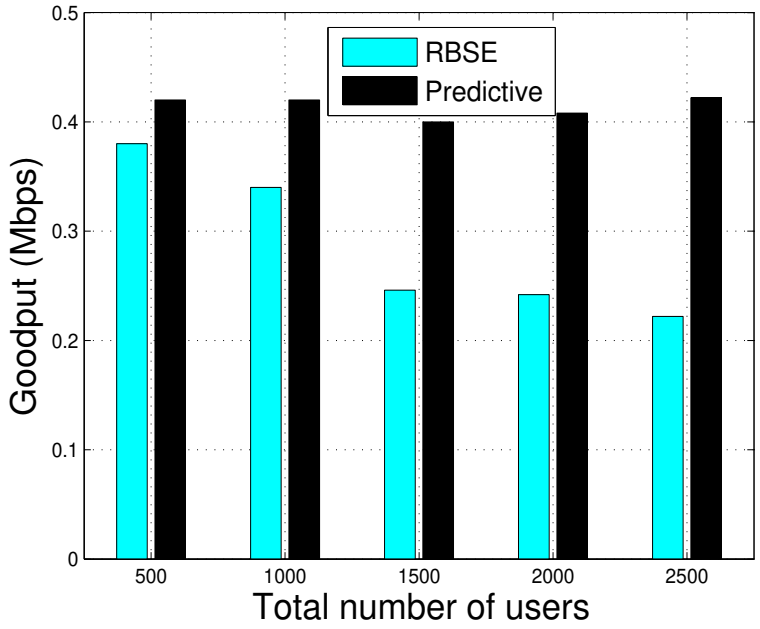


Figure 4.2: Goodput vs. traffic load

4.3.2 Simulation setup

We have considered a simulation environment similar to that considered in [35] and [39]. We consider that a macrocell LTE eNB is providing ubiquitous coverage in a 300 $meter$ $\times$ 300 $meter$ area. The total number of subcarriers of the macrocell eNB is set to 1024 [15]. The macrocell is assumed to operate in frequency division duplex mode at 2.12 GHz frequency [35]. Within the coverage region of the macrocell, 20 IEEE 802.11n femtocell APs and 10 LTE-U picocell eNBs are deployed in an un-coordinated manner. Each of the picocell eNBs are assumed to have 10 channels operating in unlicensed band where 3 channels are operating in 2.4 GHz band and 7 channels are operating in 5

GHz band [39]. All these channels are assumed to have Raleigh fading characteristics. The APs are assumed to run proportional fair MAC access mechanism. The transmission powers of the macrocell, picocells and femtocells are set to 35 dBm, 23 dBm and 13 dBm respectively [35]. The data rate at which an MT can associate with an AP has been derived based on RSSI values [78]. We consider that MTs are roaming according to smooth random waypoint mobility model [78] where the terminal velocity can vary from 3 Km/h (pedestrian) to 100 Km/h (high mobile), and acceleration ranges from 0 m/s^2 to 20 m/s^2. The slope estimator window ω' has been estimated depending on MT velocity where the sampling interval has been set to 0.01 sec [78]. Pathloss (in dB) at distance R meters from macrocell and picocells have been computed as $15.6+35\log(R)$ (outdoor environment) and from femtocells as $38.46 + 20\log(R)$ (indoor environment). Here MTs are assumed to have a strict data rate requirement for different Youtube video standards. The considered data rates are 1.5 Mbps (YouFlash), 2 Mbps (YouHtml), 2.5 Mbps (YouHD) and 2.7 Mbps (YouMob) [88]. An MT can have data rate request for any one of these video types with equal probability. We consider that the *constant bit rate* video traffic arrival follows Poisson distribution with arrival rate $6 \sim 16$ where the average holding time is exponentially distributed with its mean normalized to unity [51]. The value of Δ has been set to 0.9. We have used the standard parameter values [19] to simulate the PMIPv6 protocol.

4.3.3 Results

Figures 4.2-4.4 depict the effect of total number of MTs on various performance evaluation metrics. We vary the total number of MTs from 500 MTs to 2500 MTs with a step of 500 MTs. The result shows that our proposed predictive handoff mechanism significantly outperforms the RBSE based handoff mechanism in terms of goodput (performance gain $6 \sim 29$ percent approximately), handover failure probability (performance gain $52 \sim 81$ percent) and number of handover (performance gain $51 \sim 75$ percent). This is because instantaneous measurements of RBSE can not capture the effect of varying link condition. The link conditions in licensed band fluctuates drastically due to the varying level of interference caused by independent subcarrier allocation in neighboring access networks. On the other hand, channels operating in unlicensed band shows an intermittent characteristics due to the interference caused by LTE-U and WLAN co-

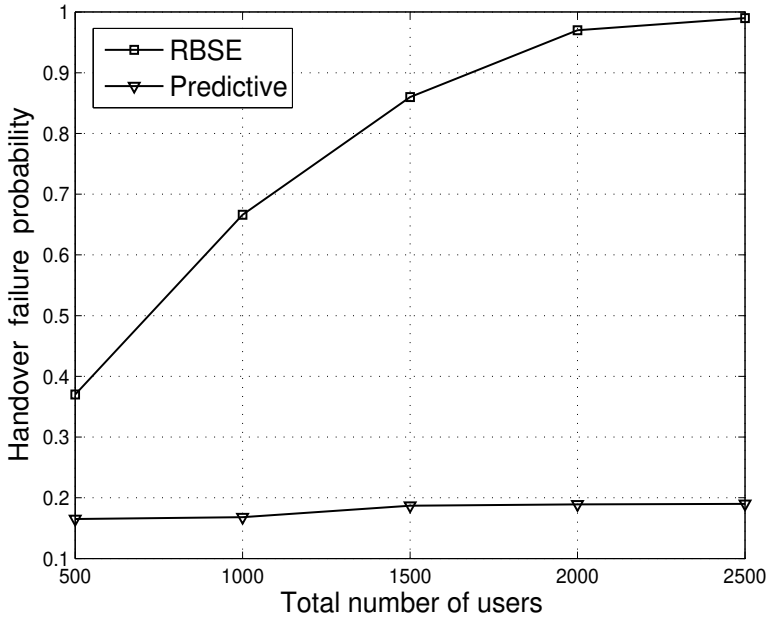


Figure 4.3: Handover failure probability vs. traffic load

existence in unlicensed band. Apart from that increased load in nearby LTE−U eNBs may cause the energy level of WLAN channels to increase beyond clear channel assessment (CCA) threshold. Consequently, some of the WLAN APs may temporarily suspend data transmission. Clearly, these effects can not be captured through the instantaneous measurements of RBSE. In contrast, our proposed predictive handover approach is based on estimated values of achievable throughput from different candidate access networks after handover execution. Moreover, the proposed approach ensures a service guarantee through the probabilistic constraints of the formulated SIP. As a result, the predictive approach significantly outperforms the RBSE based approach.

4.4 Conclusions

We have proposed a SIP based predictive handover mechanism to deal with the handover anomaly problem in 5G UDN scenario. The SIP formulation is based on the predicted throughput values from different candidate access networks after handover execution. The SIP ensures a service guarantee by imposing a lower bound on the target

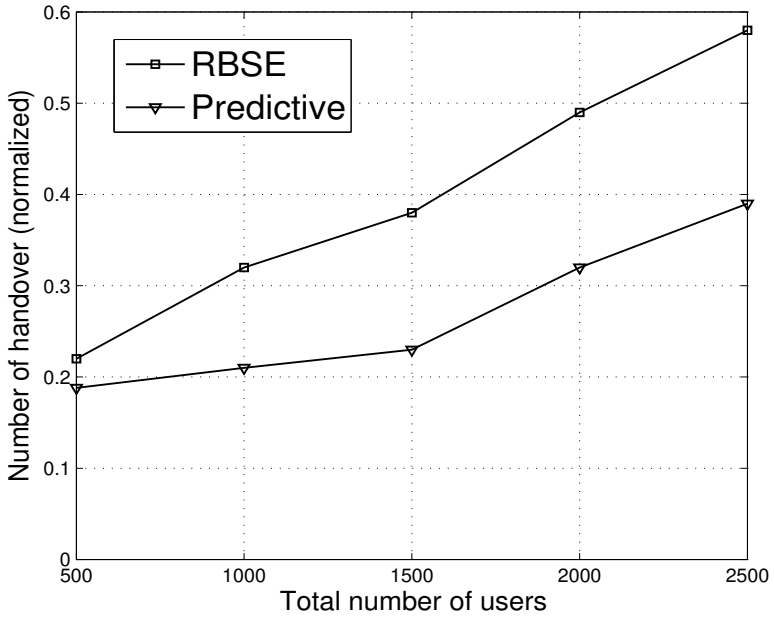


Figure 4.4: Number of handover vs. traffic load

overflow probability. To solve the SIP, we have converted the probabilistic constraint to its equivalent deterministic one and then used standard techniques involving deterministic constraints only. Extensive simulation results confirm the superiority of the proposed mechanism over RBSE based mechanism.

Appendix

The Meijer's G function $G^{p,1}_{q+1,p}\left(1, b_1, b_2, \ldots, b_q; a_1, a_2, \ldots, a_p; -z^{-1}\right)$ can be computed using the generalized hypergeometric function ${}_pF_q\left(a_1, a_2, \ldots, a_p; 1, b_1, b_2, \ldots, b_q; z\right)$ as follows [97]:

$$
\begin{aligned}
& G_{q+1,p}^{p,1}\left(1, b_1, b_2, \ldots, b_q; a_1, a_2, \ldots, a_p; -z^{-1}\right) \\
= & \frac{\prod_{j=1}^{p}\Gamma(a_j)}{\prod_{i=1}^{q}\Gamma(b_i)} \times_p F_q\left(a_1, a_2, \ldots, a_p; 1, b_1, b_2, \ldots, b_q; z\right)
\end{aligned}
$$

here p and q are integers, $\Gamma(.)$ denotes the gamma function. Now a closed form expression for ${}_pF_q\left(a_1, a_2, \ldots, a_p, m+1; b_1, b_2, \ldots, b_q, n+1; z\right)$ has been derived in [98] which is presented below:

$$
\begin{aligned}
& {}_pF_p\left(a_1, a_2, \ldots, a_p, m+1; b_1, b_2, \ldots, b_q, n+1; z\right) \\
= & \; n!\left(\frac{n-1}{m}\right)\frac{(-1)^{n(p-q)+m}}{z^n} \times \frac{(1-b_1)_n \ldots (1-b_{q-1})_n}{(1-a_1)_n \ldots (1-a_{p-1})_n} \\
\times & \left\{\sum_{k=0}^{m}\frac{(-m)_k(a_1-n)_k \ldots (a_{p-1}-n)_k}{(1-n)_k(b_1-n)_k \ldots (b_{q-1}-n)_k} \times S_1 - S_2\right\}
\end{aligned}
$$

where $S_1 =_{p-1} F_{q-1}\left(a_1-n+k, \ldots, a_{p-1}-n+k; b_1-n+k, \ldots, b_{q-1}-n+k; z\right)\frac{(-z)^k}{k!}$, $S_2 =_p F_q\left(a_1-n, \ldots, -n+1+m; b_1-n, \ldots, -n+1; z\right)$, $\{a_k\}_{k=1}^{p-1} \notin \{1, 2, \ldots n\}$, $\{b_k\}_{k=1}^{q-1} \notin \{n, n-1, \ldots n-m+1\}$ and $\frac{(m+1)_k}{(n+1)_k} = \frac{n!}{m!}\frac{1}{(k+m+1)(k+m+2)\ldots(k+n)}$. In our case, $p=3$, $q=1$, m and n are natural numbers.

Chapter 5

Performance Analysis of Dual Connectivity in Control/User-plane Split Heterogeneous Networks

Frequent handovers in HetNet cause increased control plane overhead and higher link failures. To reduce such control overhead and to ensure seamless mobility, the concept of logical separation between control plane and data plane has been evolved [4, 99]. In control/user plane (C/U) split network architecture, macrocell eNBs provide control coverage using a low frequency band signal and support efficient radio resource control (RRC) procedures for the MTs. Within the footprint of the macrocells, several small cells provide high data rate transmissions to the MTs over high frequency band signals. In C/U split network architecture, macrocells and small cells are commonly referred to as control base stations (CBSs) and data base stations (DBSs) respectively. The concept of C/U split architecture has got enough attention in the context of LTE cellular networks [65, 100]. In C/U split LTE HetNet, control plane signaling is performed by the CBS through a single link, whereas user plane data is transmitted to the MTs by the deployed DBSs. Here, DBSs do not consume radio resources from the CBS.

To deal with the high data rate demand of the forthcoming 5G networks, recently DC technology has been proposed for long term evolution (LTE) networks in Release 12. However, analyzing the performance of DC in control/user-plane (C/U) split LTE HetNet is quite limited in the preceding literature. In this chapter, an analytical frame-

work has been proposed to evaluate the performance of DC in C/U split architecture. The proposed framework explicitly considers the data rate demands of the MTs, traffic arrival pattern and channel conditions. The analytical results have also been validated against simulation results. Our analyses reveal that the performance gain of DC over traditional hard handover is actually *conditional* on underlying traffic load density and call arrival rates.

The rest of this chapter is organized as follows. In section 5.1, we describe the considered scenario. In section 5.2, we analyze the performances of DC and traditional hard handovers in C/U split network architecture. In section 5.3, we describe the results obtained from our proposed analytical framework. Finally, section 5.4 concludes the chapter. Important notations used in this chapter are depicted in Table 5.1.

Table 5.1: Important notations used to analyze the performance of DC in C/U-split HetNet

Notation	Meaning
r_j^{req}	Data rate requested by MT j.
$\Delta\left(r_j^{req}\right)$	The service coverage of MT j.
ρ_{ij}	$\frac{E_b}{N_o}$ received by MT j from DBS i through the allocated RB.
$\rho\left(r_j^{req}\right)$	The target $\frac{E_b}{N_o}$ threshold to get r_j^{req}.
B	The bandwidth of an RB.
P_{ij}	Power received by MT j from DBS i through the allocated RB.
I_{ij}	The inter-cell interference received by MT j from all neighboring DBSs of DBS i.
P_{col}	The sub-carrier collision probability.
$N(i,j)$	The set of all neighboring DBSs of DBS i from which MT j receives interference.
M	Average number of active MTs within the coverage region of a DBS.
u	Average call arrival rate of CBR traffic at any DBS.
U	Total number of RBs in a DBS.
d_{ij}	Distance between DBS i and MT j.
n	Path loss exponent.
P_{dc}	Probability that an MT can be served in DC mode.
P_{bs1}	Probability that the MT can be served only by master cell.
P_{bs2}	Probability that the MT can be served only by secondary cell.
$P_{overlap}$	Probability that the MT is residing in the overlapping region.
λ_j	Call arrival rate at MT j.

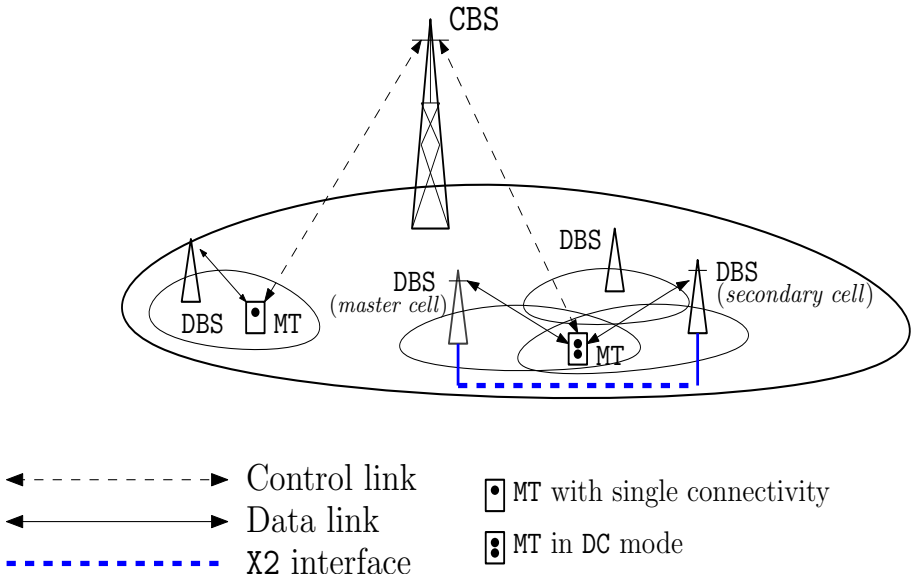


Figure 5.1: C/U split network architecture

5.1 Considered scenario

We consider a C/U split LTE HetNet where control plane and data plane are logically separated as shown in Figure 5.1. Here, a control base station (CBS) is providing ubiquitous coverage and support the necessary radio resource control (RRC) procedures. Within the coverage region of the CBS, several data base stations (DBSs) are randomly deployed. The DBSs provide high data rates over small coverage regions within the footprint of the CBS. Here, MTs are assumed to be *uniformly* distributed and moving according to random way point (RWP) mobility model [101]. Each MT is also assumed to has a strict data rate requirement for CBR video traffic. We consider that MT j is connected to the system only if the system is providing r_j^{req}, i.e., the data rate requested by MT j.

In the considered scenario, an MT can be served in either hard or DC mode at the cell edges. In hard handover mode, the MT can be served by only one DBS. An MT is served in DC mode if the requested data rate can not be served by only one DBS. In that case, the concerned MT is served by two DBSs, one of which is called *master cell* and another as *secondary cell*. The DC mechanism enables an MT to simultaneously receive data from two DBSs operating on different carrier frequencies [5]. In the considered scenario, we

adopt the 3C standard of DC where flow split occurs at the master cell and a single packet data convergence protocol (PDCP) layer is located at the master cell [17]. However, the master cell and secondary cell have two independent radio link control (RLC) layers. In this architecture, an MT in DC mode can utilize resources across both master cell and secondary cell for the same bearer. This increases the per-user throughput for a given application, which is one of the important requirements for upcoming 5G networks. Based on this scenario, in the next section, we analyze the performances of hard and DC mechanisms.

5.2 Proposed analytical framework

In this section, firstly, we model the service coverage $\Delta\left(r_j^{req}\right)$ as the lower bound on the expected distance upto which MT j gets the requested data rate r_j^{req}. Then, using the expression for $\Delta\left(r_j^{req}\right)$, we derive the expression for P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$ representing the coverage probability that an MT can be served in DC mode, the probability that the MT can be served only by the current DBS namely BS 1, the probability that the MT can be served only by the target DBS namely BS 2, and the probability that the MT which is residing in the overlapping region can be served by both BS 1 and BS 2 respectively. Finally, based on these expressions, we formulate the user association problems for DC and traditional hard handover as ILPs maximizing system throughput. Since, call drop primarily occurs due to scarcity of resource blocks (RBs) and the total demand for RBs is highly variable because of dynamic factors such as call arrival rate, channel fading and network traffic load, we analyze the upper bound on the probability that the total demand for RBs exceeds the total capacity of the serving DBSs. We denote this probability as saturation probability. In this analysis, we consider system throughput and saturation probability as performance evaluation metrics. Here, system throughput is defined as the total amount of data communicated per unit time by the system to the MTs and saturation probability is defined as the probability that the total demand imposed by the MTs exceeds the system capacity.

5.2.1 Analyzing *service coverage* $\Delta\left(r_j^{req}\right)$

In preceding literature [73, 74], service coverage regions have been widely modeled as *circular regions* mainly based on *Euclidean distance* and received signal strength (RSS) of an MT from its serving access point. This is because the throughput perceived by an MT is highly correlated with Euclidean distance and RSS values. In [73], the service coverage region of an access point has been modeled as a circular region for handover performance evaluation of voice over IP (VoIP) application. In [74], the wireless local area network (WLAN) usage efficiency has been investigated assuming circular service coverage region of WLAN access points. In practice, the area of service coverage regions are much smaller compared to that of considered in circular modeling. This is because the circular models do not consider the radio channel conditions as well as service demands of the MTs adequately. In this analysis, we explicitly consider the effect of *log normal shadowing*, *Rayleigh fading*, *network traffic load*, *call arrival rate*, *data rate requests* of the MTs as well as *inter-cell interference* from neighboring access networks to model $\Delta\left(r_j^{req}\right)$. Existing analyses for conventional cellular networks [43, 102] consider distributions of signal to interference plus noise ratio (SINR) for both data and control channels while analyzing the service coverage regions. In contrast, in our analyses, we focus only on data channel coverage while modeling $\Delta\left(r_j^{req}\right)$. This is because, during U-plane handover in C/U split architecture, C-plane is kept connected all the time and thereby saving the otherwise intensive C-plane handover signaling [44]. Consequently, RRC connection failure due to frequent handovers has no significant effect on allocating resources to the users residing in overlapping region.

According to the specification of LTE systems [15], MT j gets r_j^{req} from DBS i if the received energy per bit compared to the spectral noise density $\left(\frac{E_b}{N_o}\right)$ through the allocated resource block (RB) is sufficient to get r_j^{req}. That is:

$$\rho_{ij} \geq \rho\left(r_j^{req}\right) \tag{5.1}$$

where ρ_{ij} is the $\frac{E_b}{N_o}$ received by MT j from DBS i through the allocated RB and $\rho\left(r_j^{req}\right)$ is the target $\frac{E_b}{N_o}$ threshold to get r_j^{req} and can be computed as [54]:

$$\rho\left(r_j^{req}\right) = \frac{B}{r_j^{req}} \times \left(2^{\frac{r_j^{req}}{B}} - 1\right). \tag{5.2}$$

Here, B is the bandwidth of an RB. Assuming that the intra-cell interference is negligible and the cumulative effect of interference is much higher than that of thermal noise as considered in [15] and [103], ρ_{ij} can be computed as:

$$\rho_{ij} = \frac{B}{r_j^{req}} \times \frac{P_{ij}}{I_{ij}} \tag{5.3}$$

where P_{ij} is the power received by MT j from DBS i through the allocated RB and I_{ij} is the inter-cell interference received by MT j from all neighboring DBSs of DBS i. Now, I_{ij} can be computed as:

$$I_{ij} = \sum_{k \in N(i,j)} P_{kj} \times P_{col} \tag{5.4}$$

where P_{col} is the sub-carrier collision probability, and $N(i,j)$ is the set of all neighboring DBSs of DBS i. Since resource allocations in adjacent cells are independent in OFDMA systems [15], [103], P_{col} can be computed as:

$$P_{col} = \frac{Mean\ allocated\ resource}{Total\ amount\ of\ resource} = \frac{M \times u}{U} \tag{5.5}$$

where M is the average number of active MTs within the coverage region of a DBS, u is the average call arrival rate of CBR traffic at any DBS, and U is the total number of RBs in a DBS. Here, we are assuming that each call is served by a single RB of the serving DBS. Therefore, $M \times u$ represents the mean demand for RBs on a DBS. Here, CBR traffic arrival is assumed to follow *Poisson* distribution with rate u $(6 \sim 16)$ similar to that of [51].

To provide the just required $\frac{E_b}{N_o}$ to satisfy r_j^{req}, ρ_{ij} should be at least $\rho\left(r_j^{req}\right)$. Assuming $\rho_{ij} = \rho\left(r_j^{req}\right)$, and using equations (5.4) and (5.5), from equation (5.3) we get:

$$\rho\left(r_j^{req}\right) = \frac{B}{r_j^{req}} \times \frac{P_{ij} \times U}{\sum\limits_{k \in N(i,j)} P_{kj} \times P_{col} \times M \times u}$$
$$\Rightarrow \frac{\sum\limits_{k \in N(i,j)} P_{kj}}{P_{ij}} = \frac{B \times U}{M \times u \times r_j^{req} \times \rho\left(r_j^{req}\right)}. \tag{5.6}$$

To compute P_{ij}, the power received by MT j from DBS i through the allocated RB, we first compute the power attenuation at the receiver end considering path loss, log normal shadowing and fast Rayleigh fading as follows [104]:

$$PL_{ij} = C' \times d_{ij}^{-n} \times 10^{\frac{G}{10}} \times \pi_{ij} \tag{5.7}$$

where PL_{ij} is the attenuation of the power transmitted from DBS i and computed at MT j; C' is the constant which depends on the parameters of the considered path loss model such as frequency of transmission, height of the base station antenna and height of the mobile station antenna; d_{ij} is the distance between DBS i and MT j; n is the path loss exponent depending on the considered path loss model; G is the Gaussian distributed random variable with zero mean and standard deviation σ such that $10^{\frac{G}{10}}$ is log normally distributed; π_{ij} represents short term instantaneous power attenuation between MT j and DBS i due to Rayleigh fading, i.e., magnitude of the received signal envelop at MT j has a Rayleigh distribution. The probability density function (*pdf*) of π_{ij} is:

$$f_\pi(x) = \frac{x}{2\eta^2} \times e^{-\frac{x}{2\eta^2}}, \quad \eta \geq 0 \tag{5.8}$$

where η^2 is the variance of received power. In the subsequent analysis, we assume that all DBSs have *equal* downlink transmission power as the performance gain of DC over traditional hard handover is maximized when the MT is served with equal data rates by the serving DBSs [5]. Such an assumption is common in many studies such as [105]. Since, the allocated power to all the MTs are equal, power received at the receiver end is *proportional* to the signal path power attenuation. Accordingly, from Equations (5.6)

and (5.7) we get:

$$\frac{\sum\limits_{k\in N(i,j)} d_{kj}^{-n} 10^{\frac{G_{kj}}{10}} \pi_{kj}}{d_{ij}^{-n} 10^{\frac{G_{ij}}{10}} \pi_{ij}} = \frac{B \times U}{M \times u_j \times r_j^{req} \times \rho\left(r_j^{req}\right)}$$
$$\Rightarrow d_{ij}^{n} = \frac{K\left(r_j^{req}\right)}{Y_{ij}} \times X_{ij} \tag{5.9}$$

where $Y_{ij} = \sum\limits_{k\in N(i,j)} d_{kj}^{-n} X_{kj}$, $X_{ij} = 10^{\frac{G_{ij}}{10}} \pi_{ij}$ and $K\left(r_j^{req}\right) = \frac{B\times U}{M\times u\times r_j^{req}\times\rho\left(r_j^{req}\right)}$. We know that $d_{kj}^{-n} X_{kj},\ k \in N(i,j)$ are independent and log-normally distributed random variables with logarithmic mean $\mu + 40\log(d_{kj})$ dB [104]. It is also well known that the distribution of a sum of independent and log normally distributed random variables has no closed form expression [106]. As a result, closed form expression for the *pdf* of d_{ij} is also not known. To obtain a lower bound on $\mathbb{E}[d_{ij}]$, we first compute the expectation of $\log(d_{ij})$ and then apply *Jenson's inequality*. The Jensen's inequality states that if R is a random variable and τ is a convex function, then $\tau(\mathbb{E}[R]) \leq \mathbb{E}\left(\tau(R)\right)$. Taking logarithm in both sides of Equation (5.9) and applying *linearity* of expectation we get:

$$\mathbb{E}[\log(d_{ij})] = \frac{\log(K\left(r_j^{req}\right))}{n} + \frac{1}{n}\left(\mathbb{E}\left[\log(X_{ij})\right] - \mathbb{E}\left[\log(Y_{ij})\right]\right). \tag{5.10}$$

It may be noted that X_{ij} is the product of two random variables representing log normal shadowing and Rayleigh fading. Assuming that the log normal shadowing and Raylegh fading components are independent, the *pdf* of X_{ij} can be approximated as [104]:

$$f_X(x) = \frac{10}{\ln 10\sqrt{2\pi}\rho x} exp\left[-\frac{\left(10\log(x)-\mu\right)^2}{2\rho^2}\right] \tag{5.11}$$

where μ and ρ are the mean and standard deviation of $\log\left(X_{ij}\right)$ respectively and can be computed as $\mu = 10\log\left(2\eta^2\right) - 2.5$ dB and $\rho = \sqrt{\sigma^2 + 5.57^2}$ dB. Since, X_{ij} is a non-negative random variable, $\mathbb{E}\left[\log(X_{ij})\right]$ is computed as $\int_0^\infty \log(x) f_X(x) dx$. Now, exponential functions of the form e^{p^2} are not *integrable* for real values of p. Hence, we

compute the value of $\mathbb{E}\left[\log(X_{ij})\right]$ using Simpson's rule for solving numerical integration. Here, error bound has been set to 0.0001.

Note that Y_{ij} is a summation of a series of independent lognormally distributed random variables. We know that it has no closed form expression. However, the resultant distribution can be well approximated by another lognormal random variable [106]. To estimate the parameters of the resultant lognormal distribution, we have employed the *Fenton-Wilkinson* method. This method has been shown to have better accuracy regarding parameter estimation of the resultant log normal distribution [104, 106]. Following the Fenton-Wilkinson method, the *pdf* of Y_{ij} can be approximated as:

$$f_Y(x) = \frac{10}{\ln 10\sqrt{2\pi}b_{ij}x} exp\left[-\frac{\left(10\log(x) - v_{ij}\right)^2}{2b_{ij}^2}\right] \tag{5.12}$$

where v_{ij} is computed as:

$$v_{ij} = 10\log\left[1.1247\eta^2\sqrt{10^{\frac{\rho^2}{10}}}\sum_{k\in N(i,j)} d_{kj}^n\right] - \frac{b_{ij}^2}{2}, \text{ where}$$

$$b_{ij}^2 = 10\log\left\{1 + \frac{\left(10^{\frac{\rho^2}{10}} - 1\right)\sum_{k\in N(i,j)} d_{kj}^{-n}}{\left(\sum_{k\in N(i,j)} d_{kj}^{-n}\right)^2}\right\}.$$

Similar to X_{ij}, Y_{ij} is a non-negative random variable. Hence, the value of $\mathbb{E}[\log(Y_{ij})]$ has been computed as $\int_0^\infty \log(x) f_Y(x) dx$. Here also we employ Simpson's rule where error bound has been set to 0.0001. Substituting the value of $\mathbb{E}[\log(X_{ij})]$ and $\mathbb{E}[\log(Y_{ij})]$ in Equation (5.10), the final expression for $\mathbb{E}[\log(d_{ij})]$ can be written as:

$$\begin{aligned}\mathbb{E}[\log(d_{ij})] &= \frac{\log(K\left(r_j^{req}\right))}{n} + \frac{1}{n}\int_0^\infty \log(x) f_X(x) dx \\ &- \frac{1}{n}\int_0^\infty \log(x) f_Y(x) dx. \end{aligned} \tag{5.13}$$

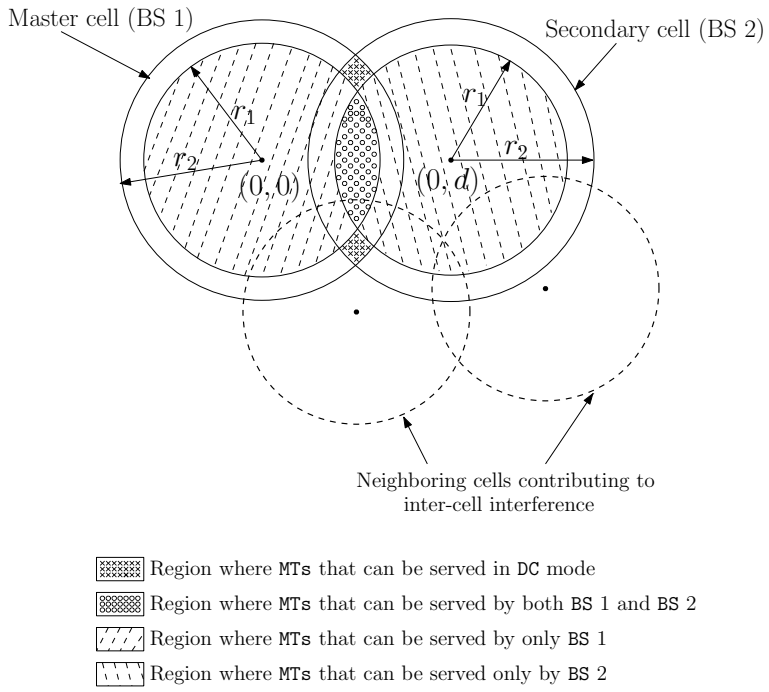


Figure 5.2: Considered model

To compute the lower bound $\Delta\left(r_j^{req}\right)$ on $\mathbb{E}\left[d_{ij}\right]$ from $\mathbb{E}\left[\log\left(d_{ij}\right)\right]$, we apply Jensen's inequality. Since 10^x is a convex function for real values of x, applying Jensen's inequality we get:

$$\Delta\left(r_j^{req}\right) \;=\; 10^{E[\log(d_{ij})]} = \frac{\left[K\left(r_j^{req}\right)\right]^{\frac{1}{n}} \times 10^{\frac{1}{n}\int_0^\infty \log(x) f_X(x) dx}}{10^{\frac{1}{n}\int_0^\infty \log(x) f_Y(x) dx}} \tag{5.14}$$

In our considered scenario, for each MT there is one serving cell and multiple neighboring cells. As per the standard [5,56], DC mechanism can be applied between the serving cell and any one of the neighboring cells where the MT can move. Here, the neighboring cell located at minimum Euclidean distance from the user is considered as the target cell. The serving cell is typically designated as *master cell* and the neighboring target cell as *secondary cell*. The DC mechanism can be applied only between this pair of cells.

In Figure 5.2, we denote the master cell as BS 1 and secondary cell as BS 2. We

model the service coverage regions of BS 1 and BS 2 by two circles each having *unit* radius by normalizing the value of $\Delta\left(r_j^{req}\right)$. We consider that BS 1 is centered at location $(0, 0)$ and BS 2 is centered at $(0, d)$ where $d < 1$. While computing the service coverage regions of BS 1 and BS 2, our model explicitly considers the inter-cell interference received from all neighboring cells. This makes our model adaptable to dense deployment scenarios in forthcoming 5G networks. Based on this model, in the next subsection, we derive the expressions for P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$.

5.2.2 Analyzing P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$

In contrast to the existing analyses [73, 74] where *circular* models are based on *Euclidean distance* and RSS values, in this work, we have considered the effect of *log normal shadow fading*, *Rayleigh fading*, *network traffic load*, *call arrival rate* and *data rate requests* of the MTs to model $\Delta\left(r_j^{req}\right)$. We denote by R_1 and R_2, the random variables representing normalized distance of an MT from BS 1 and BS 2 respectively. Further, r_1 and r_2 denote the normalized value of $\Delta\left(r_j^{req}\right)$ and $\Delta\left(r_j^{req}/2\right)$ respectively. Here, $\Delta\left(r_j^{req}/2\right)$ is the lower bound on the expected distance upto which an MT gets $r_j^{req}/2$ rate and can be obtained from Equation (5.14). In this analysis, we explicitly consider the effect of *terminal mobility*. Now, for an MT traveling under RWP mobility model within a circle of unit radius, the *pdf* of the distance of the MT from center of a circle can be approximated using *polynomial approximation* method as [73, 107]:

$$f_R(x) = \frac{6x}{257}(1 - x^2)(189 - 44x^2 - 18x^4) \tag{5.15}$$

In our considered system model, R_1 and R_2 are two *identically distributed* random variables with *pdf* $f_R(x)$. Assuming that R_1 and R_2 are independent random variables, $f_{R_1R_2}(x, y)$, the *joint cumulative probability distribution function* of R_1 and R_2 can be expressed as the product of individual probability density functions of R_1 and R_2 as follows:

$$f_{R_1R_2}(x, y) = f_{R_1}(x)\, f_{R_2}(y) \tag{5.16}$$

Here, f_{R_1} and f_{R_2} are *pdfs* of R_1 and R_2 respectively and are same as f_R.

In this analysis, we consider that, an MT is served in DC mode when the MT is getting the requested data rate neither from BS 1 nor from BS 2, but the cumulative data rate

received from the serving base stations satisfy the requested data rate. More formally, an MT is served in DC mode when both the events $\{r_1 \leq R_1 \leq r_2\}$ and $\{r_1 \leq R_2 \leq r_2\}$ occur simultaneously (depicted in Figure 5.2). The performance gain of DC over traditional hard handover is maximized when the MT is served with equal data rates by the serving DBSs [5]. Following the fact that R_1 and R_2 are independent and identically distributed random variables, the expression for P_{dc} can be computed as:

$$\begin{aligned} P_{dc} &= P\{r_1 \leq R_1 \leq r_2, r_1 \leq R_2 \leq r_2\} \\ &= \int_{r_1}^{r_2}\int_{r_1}^{r_2} f_{R_1R_2}(x,y)dxdy \\ &= \int_{r_1}^{r_2} f_{R_1}(x)\,dx \int_{r_1}^{r_2} f_{R_2}(y)\,dy \\ &= \left[\int_{r_1}^{r_2} f_R(x)\,dx\right]^2 \\ &= 4.84t^4 - 5.94t^6 + 1.38t^8 + 0.48t^{10} \\ &\quad -0.11t^{12} - 0.01t^{14} + 0.0025t^{16}. \end{aligned} \tag{5.17}$$

where $t^i = r_2^i - r_1^i$, $i \in \{4, 6, 8, 10, 12, 14, 16\}$.

The MT can be served only by BS 1 if the MT is residing within the coverage region of BS 1 but outside the service coverage region of BS 2, i.e., both the events $\{R_1 \leq r_1\}$ and $\{R_2 > r_1\}$ occur simultaneously. Hence, the expression for P_{bs1} can be derived as

follows:

$$\begin{aligned}
P_{bs1} &= P\{R_1 \leq r_1, R_2 > r_1\} \\
&= P(R_1 \leq r_1) - P(R_1 \leq r_1, R_2 \leq r_1) \\
&= \int_0^{r_1} f_{R_1}(x)dx - \int_0^{r_1}\int_0^{r_1} f_{R_1R_2}(x,y)dxdy \\
&= \int_0^{r_1} f_{R_1}(x)dx - \int_0^{r_1} f_{R_1}(x)dx \int_0^{r_1} f_{R_2}(y)dy \\
&= \int_0^{r_1} f_R(x)dx - \left[\int_0^{r_1} f_R(x)dx\right]^2 \\
&= 2.20r_1^2 - 7.55r_1^4 + 12.02r_1^6 - 7.73r_1^8 - 0.322r_1^{10} \\
&- 0.26r_1^{12} + 0.01r_1^{14} + 0.0023r_1^{16}.
\end{aligned} \tag{5.18}$$

Similar to P_{bs1}, we derive the expression for P_{bs2}. The MT can be served only by BS 2 if the MT is residing outside the service coverage region of BS 1 and within the coverage region of BS 2, i.e., both the events $\{R_1 > r_1\}$ and $\{R_2 \leq r_1\}$ occur simultaneously. Hence, the expression for P_{bs2} can be derived as follows:

$$\begin{aligned}
P_{bs2} &= P\{R_1 > r_1, R_2 \leq r_1\} \\
&= P(R_2 \leq r_1) - P(R_1 \leq r_1, R_2 \leq r_1) \\
&= \int_0^{r_1} f_{R_2}(x)dx - \int_0^{r_1}\int_0^{r_1} f_{R_1R_2}(x,y)dxdy \\
&= \int_0^{r_1} f_{R_2}(x)dx - \int_0^{r_1} f_{R_1}(x)dx \int_0^{r_1} f_{R_2}(y)dy \\
&= \int_0^{r_1} f_R(x)dx - \left[\int_0^{r_1} f_R(x)dx\right]^2 \\
&= 2.20r_1^2 - 7.55r_1^4 + 12.02r_1^6 - 7.73r_1^8 - 0.322r_1^{10} \\
&- 0.26r_1^{12} + 0.01r_1^{14} + 0.0023r_1^{16}.
\end{aligned} \tag{5.19}$$

The MT can be served by both BS 1 and BS 2 if the MT is residing within the overlapping region of service coverages of both BS 1 and BS 2, i.e., both the events $\{R_1 \leq r_1\}$ and

$\{R_2 \leq r_1\}$ occurs simultaneously. Hence, the expression for $P_{overlap}$ can be computed as follows:

$$\begin{aligned} P_{overlap} &= P\{R_1 \leq r_1, R_2 \leq r_1\} \\ &= \int_0^{r_1}\int_0^{r_1} f_{R_1R_2}(x,y)dxdy \\ &= \int_0^{r_1} f_{R_1}(x)dx \int_0^{r_1} f_{R_2}(y)dy \\ &= \left[\int_0^{r_1} f_R(x)dx\right]^2 \\ &= 4.84r_1^4 - 11.92r_1^6 + 7.78r_1^8 - 0.322r_1^{10} \\ &- 0.26r_1^{12} + 0.01r_1^{14} + 0.0025r_1^{16}. \end{aligned} \tag{5.20}$$

It may be noted that during U-plane handover in C/U split architecture, C-plane is kept connected all the time and thereby saving the otherwise intensive C-plane handover signaling [44]. Due to such feature, RRC connection failure has no significant effect on P_{dc} and $P_{overlap}$.

In subsequent subsection, based on P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$, we formulate the user association problem explicitly considering the possibility of dual connectivity. It may be noted that the traditional user association problem which is well known as NP-hard [108], refers to the problem of assigning MTs to the serving base stations without considering the possibility of dual connection.

5.2.3 Analyzing throughput performance of DC and hard handover

In this subsection, we formulate the user association problem as ILPs when DC and hard handover mechanisms are used. The *objectives* of the formulated ILPs are to maximize system throughput subject to the *resource constraints* that total demands for RBs incurred by the MTs residing within the service coverage regions of BS 1 and BS 2 do not exceed the individual capacity of the serving DBSs.

5.2.3.1 Performance model for DC

To formulate the ILPs, we first determine A, the set of MTs that can be served in DC mode, B, the set of MTs that can be served by only BS 1, C, the set of MTs that can be served by only BS 2 and D, the set of MTs that can be served by any one of BS 1 and BS 2 in traditional hard handover mode without using DC facility. It may be noted that A, B, C and D are mutually disjoint sets. Since, the MTs are uniformly distributed, the cardinality of A, B, C and D can be computed by multiplying M with P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$ respectively, and finally considering the ceiling values. That is $|A| = \lceil MP_{dc} \rceil$, $|B| = \lceil MP_{bs1} \rceil$, $|C| = \lceil MP_{bs2} \rceil$ and $|D| = \lceil MP_{overlap} \rceil$. Clearly, these cardinalities are integer values and explicitly depends on r_1 and r_2. Without loss of generality, we define the index sets of A, B, C and D as $J_A = \{1, \ldots, |A|\}$, $J_B = \{|A|+1, \ldots, |A|+|B|\}$, $J_C = \{|A|+|B|+1, \ldots, |A|+|B|+|C|\}$ and
$J_D = \{|A|+|B|+|C|+1, \ldots, |A|+|B|+|C|+|D|\}$ respectively. We also define the set $J = J_A \cup J_B \cup J_C \cup J_D$. To formulate the ILP, we define the following set of binary variables.

$$x_i^j = \begin{cases} 1 & \text{if MT } i \text{ is selected for association with BS } j. \\ 0 & \text{otherwise.} \end{cases}$$

where $i \in J$ and $j \in \{1, 2\}$. It may be noted that $x_i^1 = 0$ for all $i \in J_C$, and $x_i^2 = 0$ for all $i \in J_B$. We denote by r_i^{req} and λ_i, the data rate request of MT i and CBR call arrival rate at MT i respectively. Here, our objective is to maximize S, the system throughput. We define system throughput as the total amount of data communicated per unit time to the MTs by the system. To serve the data rate request of an MT belonging to set A, both BS 1 and BS 2 need to allocate one RB to the co-ordinating MT, i.e., 2 RBs are needed to serve a call in DC mode. In that case, each of the base stations serves half of the data rate request. On the other hand, to serve an MT belonging to sets B, C or D, one RB is sufficient. Assuming CBR traffic arrival, the objective function S can be represented as:

$$S = \sum_{i \in J_B \cup J_D} \lambda_i r_i^{req} x_i^1 + \sum_{i \in J_C \cup J_D} \lambda_i r_i^{req} x_i^2 + \frac{1}{2} \sum_{i \in J_A} \lambda_i r_i^{req} \left(x_i^1 + x_i^2 \right). \tag{5.21}$$

Next, we introduce a binary variable δ_i to indicate whether MT i is served in DC mode, i.e., $\delta_i = 1$ if MT i is served in DC mode, and $\delta_i = 0$ otherwise. Clearly, $\delta_i = 0$ for all MTs belonging to sets B, C and D. For all MTs belonging to set A, the following constraint must hold:

$$x_i^1 + x_i^2 = 2\delta_i, \forall i \in J_A. \tag{5.22}$$

To serve a call of an MT belonging to set D, only one RB need to be allocated either by BS 1 or by BS 2. Hence, the following constraint must hold:

$$x_i^1 + x_i^2 \leq 1, \forall i \in J_D. \tag{5.23}$$

Now, the MTs belonging to set B can be served by only BS 1. On the other hand, MTs belonging to set A and set D also consume resources from BS 1. In order to avoid call drop, the total demand for RBs incurred by the MTs belonging to sets A, B and D should be less than the capacity of BS 1. That is the following constraint must hold:

$$\sum_{i \in J_A \cup J_B \cup J_D} \lambda_i x_i^1 \leq U. \tag{5.24}$$

Similarly, the total demand for RBs incurred by the MTs belonging to sets A, C and D should be less than the capacity of BS 2. That is the following constraint must hold:

$$\sum_{i \in J_A \cup J_C \cup J_D} \lambda_i x_i^2 \leq U. \tag{5.25}$$

Hence, the user association problem for DC mechanism can be formulated as an ILP as follows.

ILP 1: Maximize S

subject to

$$\sum_{i \in J_A \cup J_B \cup J_D} \lambda_i x_i^1 \leq U$$

$$\sum_{i \in J_A \cup J_C \cup J_D} \lambda_i x_i^2 \leq U$$

$$x_i^1 + x_i^2 = 2\delta_i, \forall i \in J_A$$

$$x_i^1 + x_i^2 \leq 1, \forall i \in J_D$$

$$x_i^1 = 0, \forall i \in J_C$$

$$x_i^2 = 0, \forall i \in J_B$$

$$\delta_i = 0, \forall i \in J_B \cup J_C \cup J_D$$

$$x_i^1 \in \{0,1\}, x_i^2 \in \{0,1\}, \delta_i \in \{0,1\}, \forall i \in J.$$

5.2.3.2 Performance model for hard handover

In hard handover mode, an MT is connected to only one DBS at any time and therefore can not be served in DC mode, i.e., $\delta_i = 0,\ \forall i \in J$. Accordingly, MTs belonging to set A can not be served. Hence, the objective function S' for hard handover mechanism can be written as follows.

$$S' = \sum_{i \in J_B \cup J_D} \lambda_i r_i^{req} x_i^1 + \sum_{i \in J_C \cup J_D} \lambda_i r_i^{req} x_i^2. \tag{5.26}$$

Here, S' denotes the system throughput in hard handover mechanism. Similar to DC mechanism, constraint (5.23) also holds for hard handover. In order to avoid call drop, the total demand for RBs incurred by the MTs belonging to sets B and D should be less than the capacity of BS 1. That is:

$$\sum_{i \in J_B \cup J_D} \lambda_i x_i^1 \leq U. \tag{5.27}$$

Similarly, the total demand for RBs incurred by the MTs belonging to sets C and D should be less than the capacity of BS 2. Hence, the following constraint:

$$\sum_{i \in J_C \cup J_D} \lambda_i x_i^2 \leq U. \tag{5.28}$$

Hence, the ILP formulation for hard handover mechanism is as follows.

ILP 2: Maximize S'

subject to

$$\sum_{i \in J_B \cup J_D} \lambda_i x_i^1 \leq U$$

$$\sum_{i \in J_C \cup J_D} \lambda_i x_i^2 \leq U$$

$$x_i^1 + x_i^2 \leq 1, \forall i \in J_D$$

$$x_i^1 = 0, \forall i \in J_A \cup J_C$$

$$x_i^2 = 0, \forall i \in J_A \cup J_B$$

$$x_i^1 \in \{0,1\}, x_i^2 \in \{0,1\}, \forall i \in J.$$

The ILP 1 has $3\left(|A| + |B| + |C| + |D|\right)$ variables and $|A| + 2|B| + 2|C| + 2|D| + 2$ constraints. The ILP 2 has $2\left(|A| + |B| + |C| + |D|\right)$ variables and $2|A| + |B| + |C| + |D| + 2$ constraints. It may be noted that ILP 2 is a special case of ILP 1 when $\delta_i = 0,\ \forall i \in J$. For both the ILPs, the numbers of variables and constraints are determined by P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$, which in turn depends on r_1 and r_2 as depicted in Equations (5.17)-(5.20).

5.2.3.3 Complexity and solution

The traditional user association problem is known to be an NP-hard problem [108]. The considered problem is a variation of the traditional one with a possibility of dual connectivity. It may be noted that, a special case of the considered problem where $A = B = C = \phi$ is the 0-1 multiple knapsack problem, which is also widely known to be NP-complete. Here, each MT belonging to set D can be considered as an *item*. Therefore, the special case of the considered problem has $|D|$ number of items. Here,

item j ($j \in J_D$) has a *reward value* r_j^{req}, i.e., the amount of throughput MT j can provide. *Weight* of each item is the number of RB required to serve the corresponding MT. Hence, the weight of item i belonging to set D is λ_i. We have two knapsacks and the maximum weight capacity of each knapsack is U. Given these capacity constraints, the goal is to assign items to knapsacks in such a way that the total reward is maximized. Thus, the considered problem boils down to the 0-1 multiple knapsack problem, which is NP-complete. We have solved the formulated ILPs using GNU linear programming kit (GLPK) [109]. The GLPK solver gives a solution of an ILP based on LP-relaxation technique.

In the next subsection, we analyze the upper bounds on saturation probability. For the sake of simplicity and tractability, in the subsequent analyses we consider the case where the call arrival rate of all the MTs are equal, i.e., $\lambda_j = \lambda, \forall\, j \in J$.

5.2.4 Analyzing upper bounds on saturation probability

In C/U split network architecture, call drop primarily occurs due to scarcity of RBs in serving DBSs. It may be noted that the total demand for RBs T is a highly variable quantity because it depends on several dynamic factors such as call arrival rate, channel fading as well as network traffic load. So, it is worthy to analyze an *upper bound* $\beta^{(.)}$ on the probability that the total demand imposed by the MTs exceeds the system capacity, i.e., $\mathbb{P}\left(T \geq 2U\right)$ where $(.)$ indicates either hard or DC mechanism. We define this probability as *saturation probability* (*sp*). It may be noted that 2 RBs are required to serve an MT belonging to set A, whereas 1 RB is required to serve an MT belonging to sets B, C or D. Assuming $\lambda_i = \lambda,\ \forall i \in J$ the expression for T can be written as:

$$T = 2\lambda \sum_{i \in J_A} \delta_i + \lambda \sum_{i \in J \setminus J_A} \left(x_i^1 + x_i^2\right) \tag{5.29}$$

To ensure an upper bound $\beta^{(.)}$ on *sp*, the following probabilistic constraint must hold:

$$\mathbb{P}\left(T \geq 2U\right) \leq \beta^{(.)}. \tag{5.30}$$

Now, using Markov's inequality we get:

$$\mathbb{P}\left(T \geq 2U\right) \leq \frac{\mathbb{E}\left[T\right]}{2U}. \tag{5.31}$$

Assuming that all MTs are equally likely of being selected and call arrival rate at all MTs are λ, $\mathbb{E}\left[T\right]$ can be computed using *linearity of expectation* as follows.

$$\begin{aligned}
\mathbb{E}\left[T\right] &= 2\lambda \sum_{i \in J_A} \mathbb{E}\left[\delta_i\right] + \lambda \sum_{j \in J \setminus J_A} \mathbb{E}\left[x_i^1 + x_i^2\right] \\
&= \frac{2\lambda|A|}{M} + \frac{\lambda\left(|B| + |C| + |D|\right)}{M} \\
&= \lambda\left(2P_{dc} + P_{bs1} + P_{bs2} + P_{overlap}\right).
\end{aligned}$$

From Equations (5.30) and (5.31), we get:

$$\beta^{(.)} \leq \frac{\mathbb{E}\left[T\right]}{2U}. \tag{5.32}$$

Putting the value of $\mathbb{E}\left[T\right]$ in Equation (5.32), the expression for β^{dc}, an upper bound on *sp* for DC mechanism can be obtained as follows:

$$\beta^{dc} \leq \frac{\lambda}{2U}\left(2P_{dc} + P_{bs1} + P_{bs2} + P_{overlap}\right). \tag{5.33}$$

It may be noted that for a given λ, U, P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$, Equation (5.33) depicts an upper bound on β^{dc}. Similarly, putting $P_{dc} = 0$ in Equation (5.33), β^{hard}, an upper bound on *sp* for hard handover mechanism can be computed as follows.

$$\beta^{hard} \leq \frac{\lambda}{2U}\left(P_{bs1} + P_{bs2} + P_{overlap}\right). \tag{5.34}$$

5.3 Results and discussions

In this section, we present and validate the results obtained from our proposed analytical framework. First, we analyze the effect of call arrival rates on both $\Delta\left(r_j^{req}\right)$ and P_{dc}. Then, we analyze system throughput under low traffic load (200 MTs), medium traffic load (1000 MTs) and high traffic load (2000 MTs) conditions. Next, through extensive system level simulations, we investigate call dropping probabilities considering differ-

ent call arrival rates and traffic load conditions. Finally, we analyze upper bounds on saturation probability (*sp*).

For the sake of tractability, our analyses incorporate several approximations. In our analyses, while computing $\Delta\left(r_j^{req}\right)$, the inter-cell interference from all neighboring networks has been approximated by a log-normally distributed random variable (Equation (5.12)). Further, while computing P_{dc}, the probability distribution function of the distance of an MT from the serving access network has been approximated using polynomial approximation method. Finally, in order to compute system throughput, solutions of the formulated ILPs have been obtained using GLPK solver. GLPK solves ILPs based on LP relaxation techniques. For a maximization problem, LP relaxation techniques typically provide an upper bound on the actual solution. The parameter values considered in our analyses are depicted in Table 5.2.

Table 5.2: Parameter settings to analyze the performance of DC

Parameter	Value	Parameter	Value
η^2 [104]	1	σ [104]	8
α [51]	0.1	λ [51]	$6 \sim 16$
U	1000	n [104]	4

Since the proposed analytical framework incorporates several approximations, we validate the analytical results against simulation results. In the next subsection, we describe our considered simulation setup in detail. Subsequently, we discuss the results.

5.3.1 Simulation setup

The considered network configuration is similar to that considered in [84]. The simulation framework proposed in [84] provides several features to evaluate the performances of different interference mitigation techniques in LTE heterogeneous networks. However, this simulator do not have adequate features to simulate different handover mechanisms such as hard and DC. For that reason, we have developed our own system level simulator to validate the results obtained from our proposed analytical framework. We have developed our system level simulator in C++ programming language. We consider that a CBS is providing ubiquitous coverage in a 250×250 square meters area. Within

the coverage region of the CBS, several DBSs are randomly deployed. Following the guidelines given in 3GPP Rel-12 [110], the minimum distance between two DBSs has been set to 20 meters. Transmitting powers of the DBSs have been set to 20 dBm [84]. The propagation models for urban and suburban areas [85] have been considered to compute the pathlosses in DBSs. Here, pathloss in DBSs has been computed as [85]: $Pathloss(dB) = 24 + 45\log_{10}(d + 20)$ where d is distance in kilometers between the DBS and the corresponding MT. Bandwidth of each DBS has been set to 10 MHz. Within the considered simulation environment, MTs are assumed to be uniformly distributed and roaming according to RWP mobility model [79]. Here, terminal velocity can vary from 3 Km/h (pedestrian) to 100 Km/h (high mobile). The CBR traffic is assumed to arrive following a *Poisson distribution* with arrival rate λ $(6 \sim 16)$, where the average holding time of the CBR traffic is *exponentially distributed* with mean normalized to unity [51]. We consider the data rates for standard video traffic (384 Kbps), CBR *Youtube* video traffic as well as digital cinema package (DCP). Data rates considered for CBR *Youtube* standards are 1.5 Mbps (YouFlash), 0.2 Mbps (YouHtml), 2.5 Mbps (YouHD) and 2.7 Mbps (YouMob) [88]. Data rate requirement for DCP has been set to 256 Mbps. Here, the DBSs are assumed to use proportional fairness (PF) [111] as their MAC access mechanism. The source code of the developed system level simulator is available at [112].

5.3.2 Results

Figure 5.3 depicts the effect of call arrival rate on normalized service coverage region r_1. Note that r_1 is the normalized value of $\Delta\left(r_j^{req}\right)$. Here, the call arrival rate varies from 7 to 11 units with a step of unity. The traffic load in the system has been set to 1000 MTs. The result shows that r_1 shows a decreasing trend with increasing call arrival rate. It may also be noted that the value of r_1 decreases with increasing data rate requests. The reasons behind are as follows. Higher call arrival rate calls for higher number of subcarrier allocation in the serving DBSs. Since, subcarrier allocation in adjacent DBSs are independent, the subcarrier collision probability increases with increasing call arrival rate. Consequently, the cell edge users suffer from higher inter-cell interference as the call arrival rate increases. As a result, the expected distance upto which an MT can get the requested data rate decreases and r_1 shows a decreasing trend with increasing call

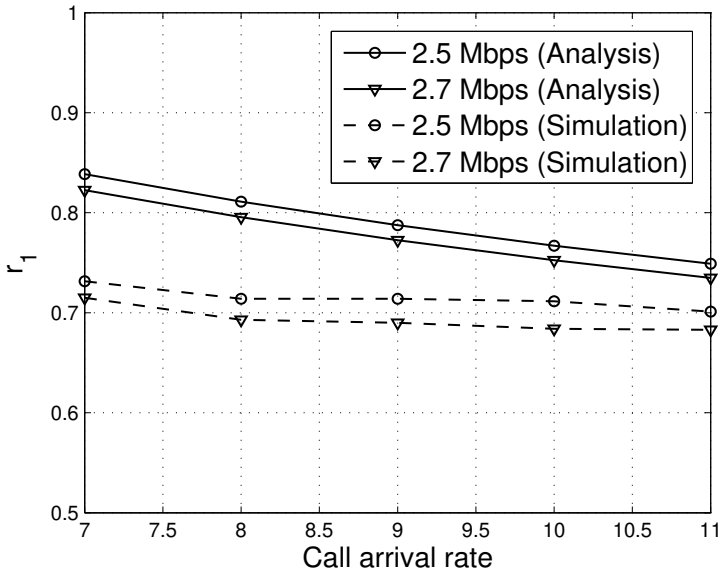


Figure 5.3: r_1 vs. call arrival rate

arrival rate for a fixed traffic load. On the other hand, for a fixed network traffic load, increasing data rate request demands for higher $\frac{E_b}{N_o}$. Since power transmitted through each subcarrier is constant, achievable $\frac{E_b}{N_o}$ by an MT is limited by pathloss, log-normal shadow fading and Raleigh fading effects. As a result, r_1 decreases as the data rate request increases.

The analytical results have also been validated against simulation results. It may be noted that while computing the service coverage region, the cumulative interference received from neighboring access networks have been modelled as a sum of a series of log normally distributed random variables. It is well known that such a summation has no closed form expression. Here, the resultant distribution has been approximated by another lognormal random variable by employing the *Fenton-Wilkinson* method (Equation (5.12)). Since, such approximation can not entirely capture the randomness of radio environment and variation of interference level, results obtained from analytical framework do not exactly match with the simulation results (maximum discrepancy 12%). However, it may be observed that both the results show a decreasing trend of r_1 with increasing call arrival rate. However, the rate of decrement is not the same because of the approximation made to arrive the analytical expressions.

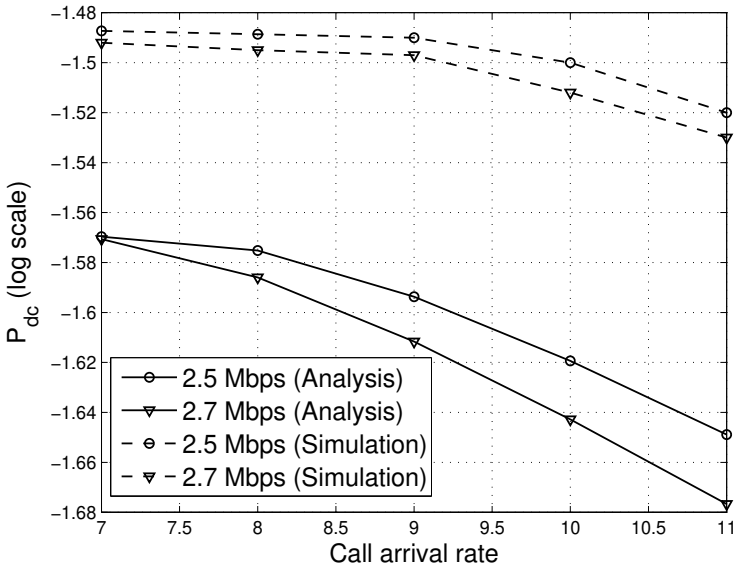


Figure 5.4: P_{dc} vs. call arrival rate

Figure 5.4 depicts the effect of call arrival rate on P_{dc}. Here, call arrival rate varies from 7 to 11 units with a step of 1 unit. The number of MTs in the system has been set to 1000 MTs. We have shown the results in log scale in order to clearly demonstrate the differences between numerical values. The result shows that P_{dc} shows a decreasing trend with increasing call arrival rate. It may also be noted that the P_{dc} value decreases with increasing data rate request. The reasons behind are as follows. An MT is served in DC mode if it can not be served by either BS 1 or BS 2 alone, but the cumulative data rate received from both the base stations satisfies the data rate request of the MT. More specifically, an MT is served in DC mode if both the events $\{r_1 \leq R_1 \leq r_2\}$ and $\{r_1 \leq R_2 \leq r_2\}$ occurs simultaneously. An increased call arrival rate demands for higher number of RBs. On the other hand, total number of RBs in BS 1 and BS 2 are constant. As a result, both r_1 and r_2 decreases with increasing call arrival rate. Consequently, probability of *jointly* occurring the events $\{r_1 \leq R_1 \leq r_2\}$ and $\{r_1 \leq R_2 \leq r_2\}$ decreases. Since, a higher data rate request requires allocation of higher number of RBs, P_{dc} value decreases as the data rate request increases. Here also, the analytical results have been validated against simulation results. It may be noted that while computing P_{dc}, the distance of an MT from an access network have been approximated using polynomial

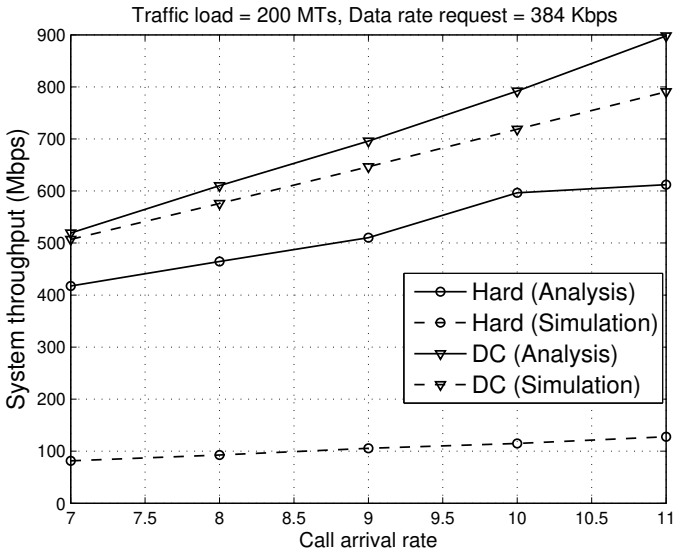


Figure 5.5: System throughput vs. call arrival rate
(under low traffic load)

approximation method (Equation (5.15)). Due to such approximation, the analytical results do not exactly match with the simulation results (maximum discrepancy of 10%). Both the analytical and simulation results show a decreasing trend of P_{dc} with increasing call arrival rate.

5.3.2.1 Comparing system throughput

In this subsection, based on our proposed analytical framework, we investigate system throughput obtained in DC mode as well as in hard handover mode. This investigation have been carried out for low traffic load (200 MTs), medium traffic load (1000 MTs) and high traffic load (2000 MTs) conditions. Data rate request considered for low, medium and high traffic load situations are 384 Kbps, 2.5 Mbps and 256 Mbps respectively. Here, the call arrival rate varies from 7 to 11 units with a step of 1 unit.

Figure 5.5 depicts the effect of call arrival rate on system throughput under lower traffic load condition. Analytical results show that the performance gain of DC over hard handover is comparatively low when the call arrival rate is less than 7 units. As the call arrival rate increases beyond 7 units, the performance gain of DC over hard handover shows an increasing trend. The reasons behind are as follows. When both the number

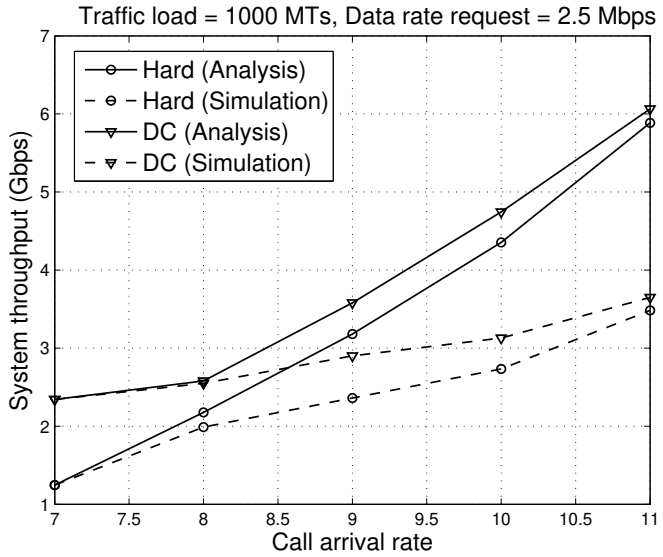


Figure 5.6: System throughput vs. call arrival rate (under medium traffic load)

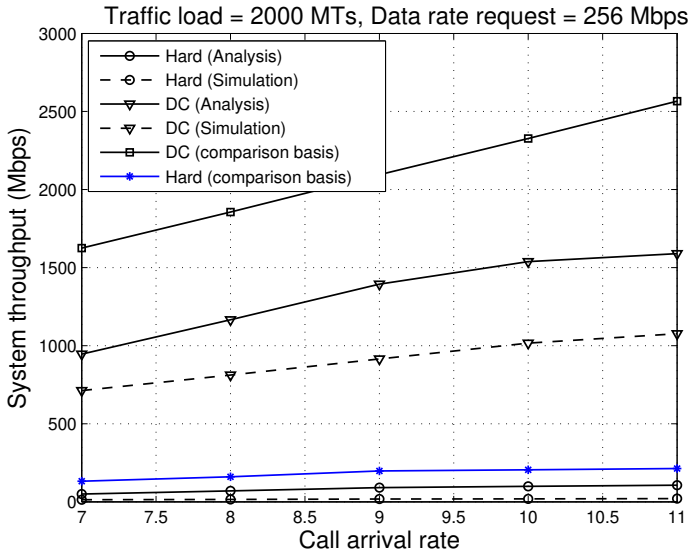


Figure 5.7: System throughput vs. call arrival rate (under high traffic load)

of MTs present in the system as well as the data rate requests are low, expected number of RBs to satisfy the data rate requests of all the MTs are also low. As a result, most of the requests can be served in hard mode. As the traffic load grows beyond 7 units, the requested number of RBs to serve the calls also increases. Consequently, the cell edge users need to be served in DC mode as their demands for RBs can be served by neither BS 1 nor BS 2. As a result, the performance gain of DC over hard shows an increasing trend.

Figure 5.6 depicts the effect of call arrival rate on system throughput under medium traffic load condition. Analytical results show that the DC mechanism outperforms hard handover in terms of system throughput. In contrast to the low traffic load situation, here the performance gain of DC over hard handover shows a decreasing trend with increasing call arrival rate. The reasons behind are as follows. Under medium traffic load situation, the probability of being served in DC mode decreases with increasing call arrival rate as shown in Figure 5.4. As a result, the MTs can not make use of the DC facility.

Figure 5.7 depicts the effect of call arrival rate on system throughput under high traffic load condition. Analytical results show that the DC mechanism significantly outperforms the hard handover mechanism in terms of system throughput. This is because, the service coverage region decreases exponentially with increasing data rate request. As a result, the number of users that can be served in hard handover mode also decreases. Consequently, most of the users are served in DC mode. It may also be noted that the system throughput obtained in DC mode shows an increasing trend when call arrival rate is less than 10. As the call arrival rate grows beyond 10, the performance gain of DC become lower. This happens due to scarcity of RBs under high call arrival rate. We have also compared the results obtained from our proposed analytical framework with that of the framework proposed in [43]. Throughput analyses in [43] are based on stochastic geometry tools which explicitly consider the effect of path loss and channel fading. However, this analysis completely ignores the effect of call arrival rate, data rate request of the users as well as capacity of the serving base stations. Due to such simplifications, the framework [43] overestimates the service coverage region compared to that of our framework which considers the effect of realistic system parameters in addition to radio channel conditions. As a result, the system throughput measured by [43] is much higher compared to that of our framework. It may be observed that, our analytical results are

much closer to the simulation results compared to that of the results obtained from [43] for both hard and DC mechanisms. This shows that our proposed framework exhibits higher accuracy compared to that of [43].

The analytical results presented in figures 5.5, 5.6 and 5.7 have been validated against simulation results. It may be noted that the analytical results presented in these figures do not exactly match with the simulation results. In Figure 5.5, the analytical results show a performance gain upto 34% (approximately) whereas simulation results show a performance gain upto 86% (approximately). In Figure 5.6, the analytical results show a performance gain of $2\% - 40\%$ approximately, whereas the simulation results show a performance gain of $4\% - 39\%$ approximately. In Figure 5.7, the analytical results show a approximate performance gain of $95\% - 98\%$, whereas the simulation results show an approximate performance gain of $94\% - 98\%$. Here, the system throughput value has been obtained by solving ILP 1 and ILP 2 using GLPK which uses LP relaxation technique to solve ILPs. For a maximization problem, LP relaxation technique typically provides an upper bound on the actual solution. Apart from that, in our analysis, the inter-cell interference from neighboring networks have been approximated by a log-normally distributed random variable. Moreover, the probability distribution function of the distance of an MT from the serving access network has been approximated using polynomial approximation method. Due to all such approximations, the analytical results do not exactly match with the simulation results. Since the inter-cell interference increases with increasing call arrival rate, the log-normally distributed random variable exhibits poor approximation for high call arrival rates. Consequently, the gap between analytical and simulation results increases with increasing call arrival rate. However, it may be observed that both the analytical and simulation results follow similar trends.

From figures 5.5, 5.6 and 5.7 we conclude that under both low and high traffic load conditions, the performance gain of DC over hard handover increases with increasing call arrival rate. On the other hand, under medium traffic load condition, the aforementioned performance gain shows a decreasing trend.

5.3.2.2 Comparing call dropping probabilities

In this section, we investigate the effect of traffic load and call arrival rate on call dropping probabilities through extensive system level simulations. Figure 5.8 depicts the

effect of traffic load on call dropping probabilities. Here, the traffic load varies from 1000 MTs to 5000 MTs with a step of 1000 MTs, whereas the call arrival rate has been set to 7 units. The results show that when the traffic load is less than a threshold (in this case 1250 MTs), the DC mechanism outperforms hard handover in terms of call dropping probabilities. As the traffic load become higher than the threshold, the hard handover outperforms the DC mechanism. The performance gain of hard handover over DC mechanism also shows an increasing trend with increasing traffic load. The reasons behind are as follows. Serving MTs in DC mode calls for resources from both BS 1 and BS 2. When the traffic load in the system is less than a threshold ($<$ 1250 MTs), demands for RBs are also low. As a result, serving the cell edge users in DC mode do not lead to resource deficiency for the MTs that can be served by only BS 1 or BS 2. Since, higher number of MTs are served in DC mode, the DC mechanism outperforms the hard handover in terms of call dropping probability when the traffic load is less than the threshold. As the traffic load increases beyond the threshold, both the demand for RBs in the system and the number of MTs that can be served in DC mode increases. Consequently, serving MTs in DC mode leads to resource scarcity for the MTs that can be served by only BS 1 or BS 2. As a result, the call dropping probability for DC mechanism is higher compared to that of hard handover when the traffic load is greater than the threshold. Since, the number of MTs that can be served in DC mode increases with increasing traffic load, the performance gain of hard handover over DC also shows an increasing trend. This observation is consistent with the results obtained in [15] where the outage probability for soft handover mechanism has been shown to be higher compared to that of hard handover due to high resource consumption. It can also be observed that for a fixed traffic load, the call dropping probability for 2.7 Mbps data rate is higher compared to that of 2.5 Mbps data rate. This is because the total demand for RBs incurred by the MTs increases with increasing data rate request as described previously.

Figure 5.9 depicts the effect of call arrival date on call dropping probabilities (shown in log scale). Here, the call arrival rate varies from 7 to 11 units with a step of 1 unit, whereas the traffic load is set to 1000 MTs. The result shows that the call dropping probability for hard handover mechanism is considerably lower compared to that of the DC mechanism for both 2.5 Mbps and 2.7 Mbps data rate. The performance gain of hard handover over DC mechanism shows an increasing trend with increasing call arrival rate.

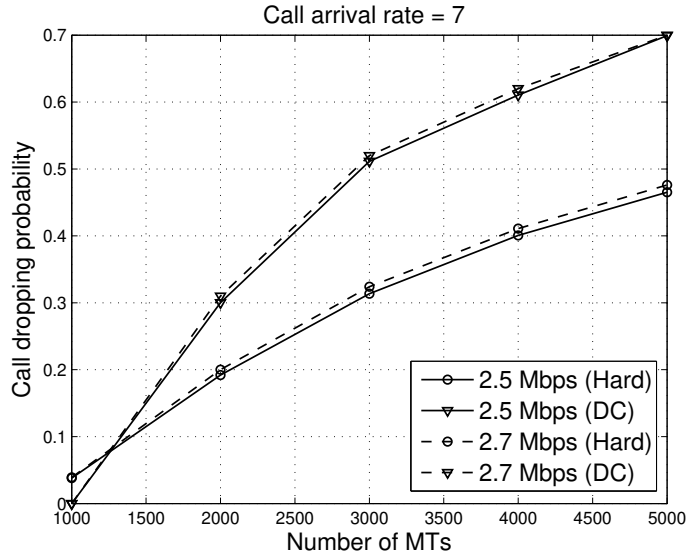


Figure 5.8: Call dropping probability vs. traffic load

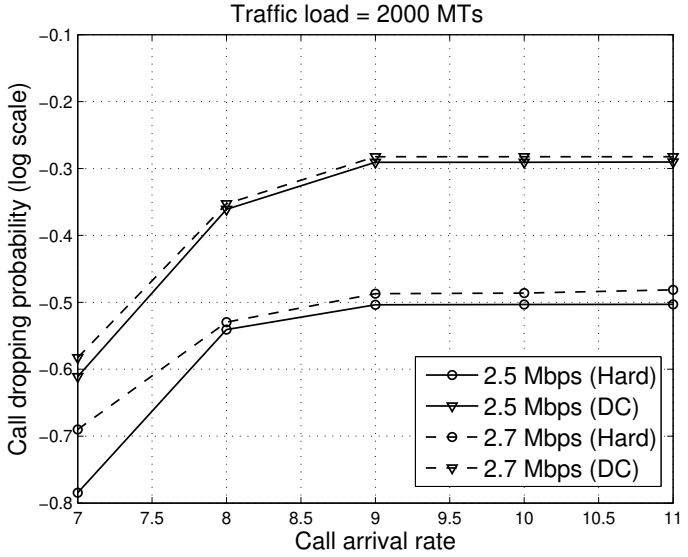


Figure 5.9: Call dropping probability vs. call arrival rate

The reasons behind are similar to that described previously.

From figures 5.8 and 5.9, we conclude that the performance gain of DC over hard handover in terms of call dropping probability is actually conditional on the underlying network traffic load. When the traffic load is less than a threshold, the call dropping probability in hard handover is higher compared to that of DC. As the traffic load increases beyond the threshold, the DC mechanism exhibits higher call dropping probability compared to that of hard handover. On the other hand, for a fixed traffic load beyond the threshold, the performance gain of hard over DC mechanism shows an increasing trend with increasing call arrival rate. It is also important to note that using DC mechanism, a higher system throughput can be achieved compared to that of hard handover (depicted in Figure 5.5 and Figure 5.6). However, DC mechanism exhibits a higher call drop when traffic load is beyond a threshold (depicted in Figure 5.8). *From these observations it is evident that the effectiveness of DC over hard handover is actually conditional on underlying network traffic load and target call dropping probabilities.*

5.3.2.3 Analyzing upper bounds on *sp*

In this section, for different data rates and traffic loads, we analyze upper bounds on *sp*, i.e., the probability that the total demand incurred by the MTs on the system exceeds total capacity. The goal of analyzing *sp* is to investigate the resource utilization efficiency under hard and DC mechanisms. Here, an increased upper bound indicates higher resource utilization.

Figure 5.10 depicts the effect of call arrival rates on upper bounds on *sp* (shown in log scale) for different data rate requests. Here, call arrival rates vary from 6 to 16 units with a step of 1 unit, whereas the traffic load has been set to 1000 MTs. The upper bounds shows an increasing trend with increasing call arrival rate for both 2.5 Mbps and 2.7 Mbps data rate requests. The reasons behind are as follows. Though an increased call arrival rate decreases the service coverage region r_1 as well as P_{dc}, the number of allocated RBs increases drastically due to high call arrival rates. Consequently, the probability of the total demand exceeding the total capacity also increases. As a result, *sp* also shows an increasing trend with increasing call arrival rate. It is also interesting to observe that for a fixed call arrival rate, the upper bound for 2.7 Mbps data rate is lower compared to that of 2.5 Mbps data rate. This is because, under high data rate demand,

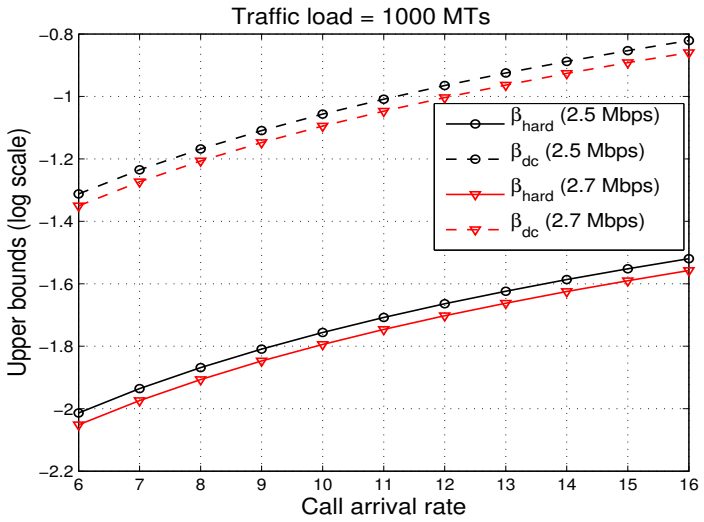


Figure 5.10: Upper bounds on saturation probability vs. call arrival rate

lesser number of MTs are served by the base stations. More specifically, for increased data rate request, the required $\frac{E_b}{N_o}$ to satisfy the data rate request also increases. As a result, the value of r_1 decreases. Due to such decrement of service coverage region, all of the probabilities P_{dc}, P_{bs1}, P_{bs2} and $P_{overlap}$ also decreases. This leads to reduced resource consumption as the number of RBs that can be allocated to an MT is fixed for both hard and DC mode. Consequently, the upper bounds show a decreasing trend for increased data rate request.

Figure 5.11 depicts the effect of traffic load on upper bounds for both 2.5 Mbps and 2.7 Mbps data rates. Here, the traffic load varies from 1000 MTs to 10000 MTs with a step of 1000 MTs. The call arrival rate has been set to 7 units. The upper bound shows a decreasing trend with increasing traffic load. The reasons behind are as follows. The service coverage region decreases drastically with increasing traffic load as can be found from Equation (5.14). A reduced service coverage region leads to reduced resource consumption. Consequently, the probability of the total demand exceeding the total capacity decreases. As a result, the upper bound shows a decreasing trend with increasing traffic load. It may also be observed that the upper bound for 2.5 Mbps data rate is higher compared to that for 2.7 Mbps data rate. This is because, for a fixed traffic load, the service coverage region shows a decreasing trend for increasing data rate request as depicted in Figure 5.3. Due to such decrement, the number of MTs that

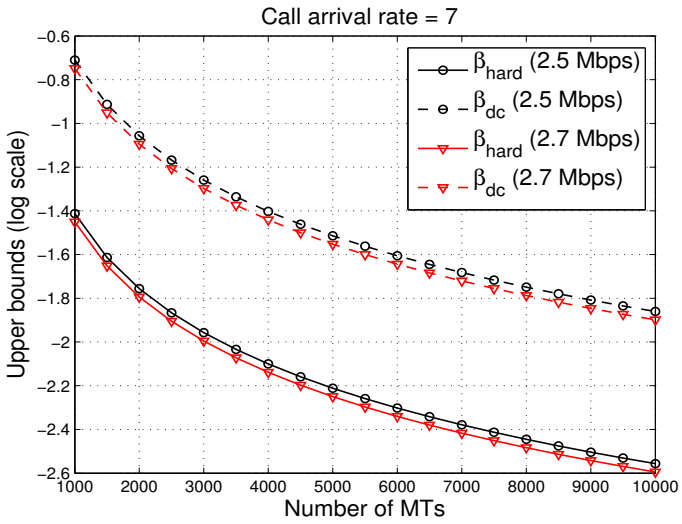


Figure 5.11: Upper bounds on saturation probability vs. Traffic load

can be served by each DBS also decreases. Consequently, the number of allocated RBs decreases with increased data rate request. As a result, the upper bound for 2.5 Mbps data rate is higher compared to that of 2.7 Mbps data rate.

From the results presented in figures 5.10 and 5.11, we conclude that the resource utilization increases with increasing call arrival rate when the traffic load is fixed. On the other hand, the resource utilization shows a decreasing trend with increasing traffic load when the call arrival rate is fixed.

5.4 Conclusion

In this chapter, we have proposed an analytical framework to compare the performances of DC and traditional hard handover in C/U split LTE heterogeneous network architecture. Our analysis explicitly considers the data rate demands of the MTs, network traffic load, call arrival rate and channel conditions. We have considered system throughput and saturation probability as performance evaluation metrics. The analytical results have also been validated against extensive simulation results. From both analyses and simulation results, we conclude the following:

- For low and high traffic load conditions, the performance gain of DC over hard

handover (in terms of system throughput) increases with increasing call arrival rate. On the other hand, for medium traffic load condition, the performance gain shows a decreasing trend.

- For both DC and hard handover mechanisms, analyzing saturation probability reveals that the resource utilization increases with increasing call arrival rate for a fixed traffic load. On the other hand, the resource utilization decreases with increasing traffic load for a fixed call arrival rate.

- The performance gain of DC over hard handover in terms of call dropping probability is actually *conditional* on the underlying network traffic load. When the traffic load is less than a threshold, the call dropping probability in hard handover is higher compared to that of DC. As the traffic load increases beyond the threshold, the DC mechanism exhibits higher call dropping probability compared to that of hard handover.

  On the other hand, for a fixed traffic load beyond the threshold, the performance gain of hard handover over DC mechanism shows an increasing trend with increasing call arrival rate.

- The effectiveness of DC over hard handover is conditional on underlying network traffic load and target call dropping probabilities.

Chapter 6

Analyzing Handover Performances of Mobility Management Protocols in Ultra Dense Networks

To deal with the exponential increase of mobile data traffic, ultra dense network (UDN) has been evolved as a promising solution for the forthcoming 5G cellular networks. In UDN, a mobile terminal (MT) experiences frequent handovers due to limited coverage regions of the deployed small cells. Such frequent handovers cause increased packet loss and blocking rate if the handover latency is very high. The handover latency explicitly depends on the layer 3 (L3) handover mechanisms of upper layer mobility management protocols (MMPs) as well as handover execution mechanisms (HEMs) operating at layer 2 (L2). The HEMs also have significant impact on L3 handover latency. Despite such dependencies, existing handover performance evaluations of MMPs do not adequately consider the effect of underlying HEMs. In this chapter, we analyze the handover performances of different class of MMPs considering the effect of underlying HEMs in terms of handover latency, handover packet loss and handover blocking rate. For analysis purpose, we consider a network layer MMP namely fast mobile IPv6, a transport layer MMP namely seamless IP diversity based generalized mobility architecture (SIGMA), and a distributed mobility management (DMM) protocol as candidate MMPs. Here hard and semisoft handovers have been considered as underlying HEMs. Our analysis reveal the conditional effect of underlying HEMs on the handover performances of upper layer

MMPs. Further, based on such analysis, we prioritize among different combinations of MMPs and HEMs employing analytic hierarchy process (AHP). Such priority assignment would serve as a protocol selector in UDN scenario.

Table 6.1: Important notations used to analyze handover latency, handover packet loss and handover blocking rate

Notation	Meaning
T_{L2}	The L2 switching delay.
T_{UNA}	Time to receive the UNA message.
T_{buff}	Time to receive the first buffered packet from the target eNB.
T_{DAD}	Delay associated with DAD which is a part of IP address configuration.
T_{FBU}	Time required to deliver the FBU message to the target eNB.
T_b	Time required for signaling through backhaul network.
$L_{ho}^{(.)}$	The handover latency for protocol $(.)$.
$B^{(.)}$	The handover blocking rate for protocol $(.)$.

The rest of the chapter is organized as follows. In section 6.1, we present the performance models for FMIPv6, SIGMA and DMM in terms of handover latency, handover packet loss and handover blocking rate. In section 6.2, we present the results obtained from our analyses. Based on these analytical results, we propose an AHP based ranking of different MMPs in section 6.3. Important notations used to develop the performance models for FMIPv6, SIGMA and DMM have been enlisted in Table 6.1.

6.1 Performance analysis

In this analysis, we have considered handover latency, handover packet loss and handover blocking rate as performance evaluation metrics. Handover latency is defined as the expected delay between disconnection of the MT from the current evolved node B (eNB) and receipt of first packet from the target eNB. Analysis of handover latency explicitly considers the effect of traffic load, data rate request, call arrival rate and HEMs. Handover packet loss is defined as the sum of all lost data packets belonging to downlink IP flows during handover. Such packet loss depends on call arrival rate, average session length as well as handover latency. Further, handover blocking rate for an MMP has been

defined as the number of handover failed per unit time due to high handover latency, signal strength deterioration at cell edges and resource scarcity at the target eNB.

In the subsequent subsections, we first describe the considered network scenario. Then, in section 6.1.2, we have computed handover latencies for FMIPv6, SIGMA and DMM when hard and semisoft handovers are used. In sections 6.1.3 and 6.1.4, we have computed handover packet loss and handover blocking rate respectively.

6.1.1 Considered network scenario

To carry out the performance evaluation, we have considered a typical HetNet architecture as depicted in Figure 6.1. Here the macrocell belongs to long term evolution advanced (LTE−A) standard. The macrocell is connected to the mobility management entity (MME) and serving gateway (SGW) through S1−C and S1−U interfaces respectively. The MME and SGW are in turn connected to the Internet backbone through evolved packet core (EPC). Within the coverage region of macrocell, several orthogonal frequency division multiple acess (OFDMA) based femtocell access points (FAPs) are randomly deployed. These FAPs may belong to different radio access technologies such as LTE−A, IEEE 802.11 wireless local area network (WLAN) and worldwide interoperability for microwave access (WiMAX). We assume that the deployed FAPs are using OFDMA technology. Hence, semisoft handover is supported by the system [15]. We consider a loose coupling [6], [90] between the macrocell and FAPs to provide unified access to the MTs within the HetNet. In loose coupling, FAPs are independently deployed and are individually connected to the Internet. FAPs belonging to LTE−A technology are connected to the SGW through S1 interface. FAPs belonging to radio access technologies other than LTE−A standard (e.g., WLAN, WiMAX) are connected to a separate gateway which in turn is connected to the Internet. Within the coverage region of macrocells, several MTs are uniformly distributed and moving according to smooth random waypoint mobility model [79]. Here, MTs are assumed to have data rate request for constant bit rate (CBR) video traffics. In the considered scenario, both macrocell and FAPs are assumed to use proportional fairness as MAC access mechanism. Due to heterogeneous architecture, we consider that mobility management is performed from upper layer by different MMPs such as FMIPv6, SIGMA and DMM.

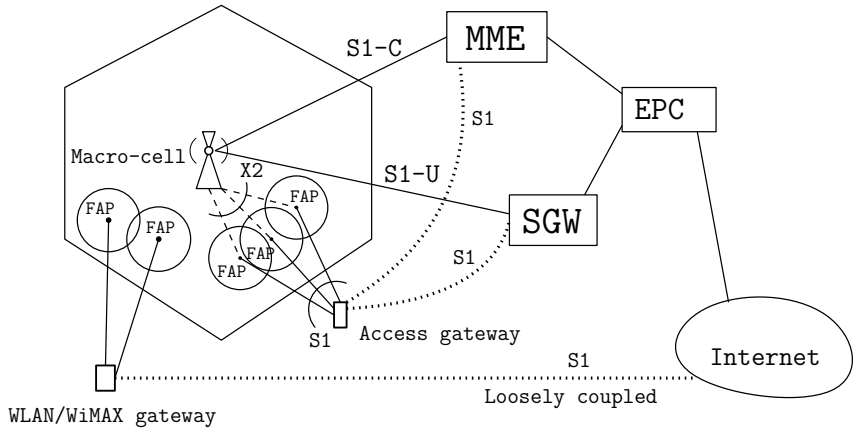


Figure 6.1: Considered HetNet scenario

6.1.2 Handover latency

FMIPv6 is a network layer MMP which can operate either in *predictive* mode or in *reactive* mode depending on situation [19]. In predictive mode operation, an MT predicts the target eNB and configures an IP address for future use while residing in the current eNB. As a result, handover latency in predictive mode operation is quite lower because IP address configuration delay is not a part of handover latency. In case of handover failure, the FMIPv6 switches to reactive mode. In this mode, an MT needs to configure the IP address after being attached to the target eNB. Consequently, IP address configuration delay becomes part of handover latency and the resulting latency becomes higher. Here we have computed handover latencies for MIPv6, SIGMA and DMM considering the effect of underlying hard and semisoft handover mechanism.

Handover latency for FMIPv6

Assuming hard handover as underlying handover execution mechanism, we first derive P^{hard}, the handover latency in predictive mode and R^{hard}, the handover latency in reactive mode. Then, based on P^{hard}, R^{hard} and handover failure probability (p), we derive the expression for F^{hard} the expected handover latency for FMIPv6 protocol when hard handover is used. In this analysis, handover failure probability (p) is defined as the probability that the target network is unable to provide the requested data rate due to high traffic load, congestion and low signal strength at the cell edges. Here we assume that p

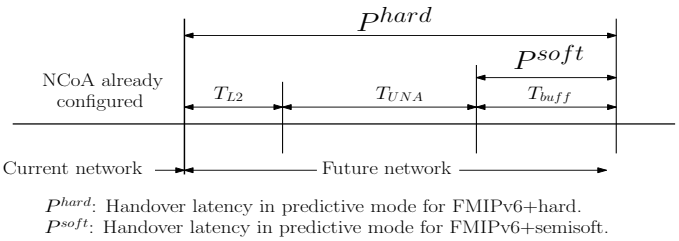


Figure 6.2: FMIPv6 handover in predictive mode operation

is known and it is computed later through simulations.

In predictive mode operation, the MT performs necessary signaling to obtain a new care of address (NCoA) for future use in target eNB while associated with current eNB. The configuration of NCoA is confirmed to the MT through fast binding acknowledgement (FBACK) message. After performing L2 switching, the MT sends an unsolicited neighbor advertisement (UNA) message to the target eNB informing the already configured NCoA. After the receipt of UNA message, the target eNB starts sending packets to the MT. *In this case L2 switching delay is a part of handover latency because of hard handover mechanism*. Denoting by T_{L2} as the L2 switching delay, by T_{UNA} the time required to receive UNA message by the new eNB from the co-ordinating MT and by T_{buff} as the time required to receive the first buffered data packet which has been forwarded from current eNB to target eNB over the X2 interface, the expression for P^{hard} can be written as:

$$P^{hard} = T_{L2} + T_{UNA} + T_{buff}. \tag{6.1}$$

Components of P^{hard} are depicted in Figure 6.2. If the target eNB cannot provide sufficient number of RBs to serve the data rate request of the MT then handover failure occurs. In case of handover failure, the target eNB may deny to configure an NCoA. Consequently, the MT cannot receive the FBACK message while associated with the current eNB. As a result, the FMIPv6 protocol switches to reactive mode. In reactive mode, after performing L2 switching, the MT requests for NCoA to the target eNB. The configuration of NCoA involves duplicate address detection (DAD) delay. After configuration of NCoA, the MT sends an UNA message and fast binding update (FBU) to the target eNB. After that the target eNB performs necessary signaling with current eNB through the backhaul network. The delay in backhaul network includes exchanges of handover

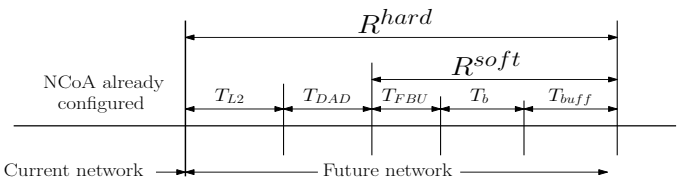
initiate (HI) and handover acknowledgement (HACK) messages, and forwarding all the buffered data packets from current eNB to the target eNB. After completion of signaling through backhaul network, the target eNB starts delivering packets to the associated MT. Denoting by T_{DAD} as the delay associated with DAD which is a part of IP address configuration, by T_{FBU} as the time required to deliver the FBU message to the target eNB, by T_b as the time required for signaling through backhaul network and by T_{buff} as the delay required to receive the first data packet from the buffered packet queue of the target eNB, the expression for R^{hard} can be written as:

$$R^{hard} = T_{L2} + T_{DAD} + T_{FBU} + T_b + T_{buff}. \tag{6.2}$$

Components of R^{hard} are depicted in Figure 6.3. Here T_b has been computed as the cumulative delay required for control message communication between current and target eNBs. Signaling delay for each control message includes processing delay, transmission delay and propagation delay at each hop and encapsulation-decapsulation delay at the tunnel end points. *It may be noted that here also L2 switching delay is a part of handover latency because of hard handover mechanism.* Based on p, P^{hard} and R^{hard} the expression for F^{hard} is computed as:

$$\begin{aligned} F^{hard} &= (1-p) \times P^{hard} + p \times R^{hard} \\ &= (1-p) \times (T_{L2} + T_{UNA} + T_{buff}) \\ &+ p \times (T_{L2} + T_{DAD} + T_{FBU} + T_b + T_{buff}) \,. \end{aligned} \tag{6.3}$$

Now, assuming semisoft handover as underlying handover execution mechanism, we



R^{hard}: Handover latency in reactive mode for FMIPv6+hard.
R^{soft}: Handover latency in reactive mode for FMIPv6+semisoft.

Figure 6.3: FMIPv6 handover in reactive mode operation

derive P^{soft}, the handover latency in predictive mode and R^{soft}, the handover latency in reactive mode. It may be noted that the effect of semisoft handover has not been considered in existing performance analysis of FMIPv6 [19], [21]. Based on P^{soft}, R^{soft} and p we derive the expression for F^{soft}, the expected handover latency for FMIPv6 protocol when semisoft handover is used. In semisoft handover, multiple control links remain active over the handover region. As a result, in predictive mode operation of FMIPv6, one control link can be utilized to maintain connectivity with the current eNB while another one can be used to send UNA message to the target eNB during handover execution. The control channel with the current eNB is released after the MT receives the first data packet from the target eNB. As a result, T_{L2} and T_{UNA} delay components are eliminated from handover latency. Hence, the expression for P^{soft}, the handover latency for predictive mode operation of FMIPv6 protocol when semisoft handover is used can be written as:

$$P^{soft} = T_{buff}. \tag{6.4}$$

Components of P^{soft} are depicted in Figure 6.2. When FMIPv6 operates in reactive mode, an MT can receive data from the current eNB when DAD mechanism is in progress. After successful configuration of NCoA, the MT releases the control channel with current eNB. As a result, T_{L2} and T_{DAD} delay components are eliminated from handover latency. Hence, the expression for R^{soft}, the handover latency for FMIPv6 protocol in reactive mode when semisoft handover is used can be written as:

$$R^{soft} = T_{FBU} + T_b + T_{buff}. \tag{6.5}$$

Components of R^{soft} are shown in Figure 6.3. Now, based on P^{soft}, R^{soft} and p, the expression for F^{soft}, the expected handover latency of FMIPv6 protocol when semisoft handover is used can be written as:

$$\begin{aligned} F^{soft} &= (1-p) \times P^{soft} + p \times R^{soft} \\ &= (1-p) \times T_{buff} + p \times (T_{FBU} + T_b + T_{buff}) . \end{aligned} \tag{6.6}$$

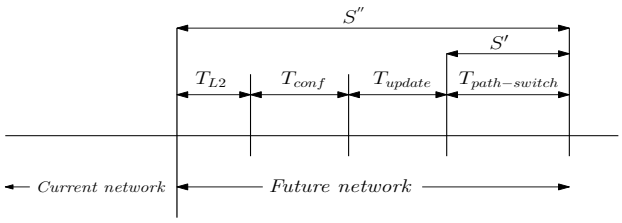


Figure 6.4: SIGMA protocol operation

Handover latency for SIGMA

In contrast to FMIPv6, SIGMA is a *transport layer* MMP which exploits the *multi-homing* and *multi-streaming* capacity of SCTP to reduce the handover latency. Handover mechanism in SIGMA consists of three phases. The *NCoA configuration phase* starts as soon as the MT moves into the overlapping region and receives agent advertisements from target eNB. Delay associated to NCoA configuration includes the delay for requesting an IP address and obtaining it through standard mechanisms of dynamic host configuration protocol (DHCP). In *NCoA update phase*, the MT notifies the correspondent node (CN) about the availability of NCoA through address dynamic reconfiguration option of SCTP. The location manager component, which acts as an anchor point for SIGMA protocol, is also updated using location update and location update ACK messages accordingly. The *path switch phase* starts as soon as MT moves further into the overlapping zone and informs the CN that the IP address corresponding to current eNB is obsolete. Accordingly, the CN sets the NCoA as the primary IP address for communicating the MT and deletes the old IP address from available list of IP address. The basic idea of SIGMA is to exclude the delay required for NCoA configuration (T_{conf}) and the delay required in NCoA update phase (T_{update}) from handover latency by exploiting multi-homing and multi-streaming capacity of SCTP protocol. However, when hard handover is used at link layer, SCTP cannot configure the NCoA in the overlapping region when data communication is in progress with current eNB. Here, handover process starts after the MT releases the control channel with current eNB. As a result, T_{L2} becomes a part of handover latency. Denoting by $T_{path-switch}$, the delay in path switch phase, S'', the handover latency for SIGMA protocol in case of successful handover, can be written as:

$$S'' = T_{L2} + T_{conf} + T_{update} + T_{path-switch}. \tag{6.7}$$

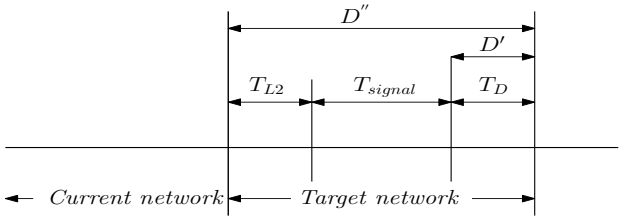


Figure 6.5: DMM protocol operation

Components of S'' are depicted in Figure 6.4. Subsequently, the expression for S^{hard}, the handover latency for SIGMA protocol when hard handover is used, can be computed as:

$$S^{hard} = \frac{1}{1-p}\left(T_{L2} + T_{conf} + T_{update} + T_{path-switch}\right). \tag{6.8}$$

In semisoft handover, the MT can simultaneously maintain control links with current as well as target eNBs over the handover region. During handover execution, control signaling corresponding to NCoA configuration and NCoA update can be performed through the control link with target eNB whereas data can be received from current eNB utilizing the multi-homing features of SCTP. As a result, T_{L2}, T_{conf} and T_{update} delay components are excluded from handover latency. Hence, the handover latency S' in case of successful handover can be written as:

$$S' = T_{path-switch} \tag{6.9}$$

However, when handover failure occurs, the MT have to experience a handover latency as long as S^{hard}. Hence, the expression for S^{soft}, the handover delay for SIGMA when used with semisoft handover, can be written as:

$$S^{soft} = (1-p) \times S'' + p \times S' \tag{6.10}$$

Components of S' and S'' are depicted in Figure 6.4.

Handover latency for DMM

The proxy MIPv6 (PMIPv6) based DMM consists of two components namely DMM gateway and control mobility database (CMD). DMM gateways act as a plain access router

to forward packets to and from the Internet. During handover, the DMM gateways can switch IP flows from current to future access network without any disruption. The CMD stores the prefixes advertised to the MTs, which DMM gateway advertised those prefixes and to which DMM gateway the MT is currently connected to. By means of extended PBU and PBA signaling, the CMD sends instruction to recover the MN's ongoing IP flows after a handover. The handover latency of DMM consists of three components: (a) L2 switching delay (T_{L2}), (b) delay associated to router solicitation, extended proxy binding update (PBU) and proxy binding acknowledgement (PBA) signaling and reception of router advertisement from the target network (T_{signal}), and (c) the delay for the first packet delivered to MT while connected to the target network (T_D). Hence, the expression for handover latency in case of successful handover can be written as:

$$D^{''} = T_{L2} + T_{signal} + T_D. \tag{6.11}$$

Components of $D^{''}$ have been depicted in Figure 6.5. Hence, the expected handover latency can be computed as:

$$D^{hard} = \frac{1}{1-p}\left(T_{L2} + T_{signal} + T_D\right). \tag{6.12}$$

If semisoft handover is used at L2, T_{L2} and T_{signal} delays are omitted from handover latency. As a result, the handover latency reduced to $D' = T_D$. Hence, the expected latency can be computed as:

$$D^{soft} = p \times D^{''} + (1-p) \times D'. \tag{6.13}$$

6.1.3 Handover packet loss

While an MT experiences its handover, data packets destined for the MT will be lost. Assuming that there is no buffer management at the network side, handover packet loss $P^{(.)}$ can be defined as the sum of all lost data packets belonging to downlink IP flows during handover. Here $(.)$ is the protocol indicator (e.g., DMM + hard). Following the formulation presented in [19], $P^{(.)}$ has been computed as:

$$P^{(.)} = \lambda_s \times L_{ho}^{(.)} \tag{6.14}$$

where λ_s is the call arrival rate and $L_{ho}^{(.)}$ is the handover latency for protocol $(.)$. It may be noted that $P^{(.)}$ is directly proportional to $L_{ho}^{(.)}$.

6.1.4 Handover blocking rate

Handover blocking rate for an MMP is defined as the number of handover failed per unit time due to high handover latency, signal strength deterioration at the cell edge and resource scarcity at the target eNB. In previous analyses of handover blocking rate [19], [113], handover latency has been considered as the principle factor without any consideration of other factors. In this analysis, first we compute the handover latency $L_{ho}^{(.)}$ considering the effect of traffic load, data rate request, call arrival rate and HEMs, and then based on $L_{ho}^{(.)}$ we compute $B^{(.)}$ the handover blocking rate. To compute the expression for $B^{(.)}$ we have adopted the formulations presented in [19], [113]. The expression for $B^{(.)}$ can be given as follows:

$$B^{(.)} = \frac{\mu \times L_{ho}^{(.)}}{1 + \mu \times L_{ho}^{(.)}}, \tag{6.15}$$

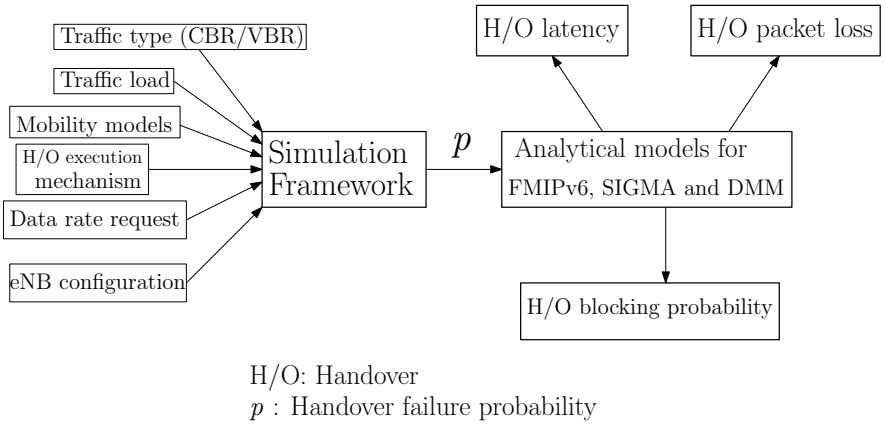
where μ is the border crossing rate for MTs and can be computed as:

$$\mu = \frac{2v}{\pi R}. \tag{6.16}$$

Here v is the average velocity of the MT and R is the radius of the radio access network.

6.2 Evaluating the combined effect

To evaluate p under different traffic load conditions, call arrival rates, data rate request and HEMs, we have developed our own system level simulator in C++. In the next subsection, first we describe the models considered in our developed system level simulator to evaluate p. Then, based on the values of p, we analyze the combined effect HEMs and MMPs on handover latency, handover packet loss and handover blocking rate. The input-output relation between simulation framework and analytical models have been demonstrated in Figure 6.6.



H/O: Handover
p : Handover failure probability

Figure 6.6: Input-output relation between simulation framework and analytical models.

6.2.1 Simulation framework to determine p

We have considered a realistic simulation environment similar to that of [84]. We consider that 7 LTE−A based macro cell eNBs (MeNBs) are providing ubiquitous coverage in a 1000×1000 m^2 area. Within the coverage region of MeNBs, multiple LTE−A based FAPs are arbitrarily deployed. Radii of MeNBs and FAPs have been set to 250 meters [84] and 100 meters [3] respectively. Transmitting powers of MeNBs and FAPs have been set to 46 dBm and 23 dBm respectively [84]. For MeNBs, the path loss have been computed using the propagation model for urban and suburban areas as described in [85]: $Pathloss(dB) = 128.1 + 37.6 \log_{10}(d)$ where d is the distance in kilometers between the MeNB and the corresponding MT. For FAPs, path loss have been computed as [85]: $Pathloss(dB) = 24 + 45 \log_{10}(d + 20)$. Total bandwidth for MeNBs and FAPs have been set to 10 MHz and 5 MHz respectively. Within the considered HetNet, MTs are assumed to be uniformly distributed and roaming according to *smooth random waypoint* mobility model [79]. Here, terminal velocity can vary from 3 Km/h (pedestrian) to 100 Km/h (high mobile), and acceleration ranges from 0 m/s^2 to 20 m/s^2. An MT accelerates or decelerates depending on its current and target velocity. Here CBR traffic is assumed to arrive following a *Poisson distribution* with arrival rate λ [51]. The average holding time of the CBR traffic is *exponentially distributed* and its mean is normalized to unity. Here MTs are assumed to have data rate requirements for *digital cinema package* (DCP) application. Considered data rate for DCP is 256 Mbps. Such a scenario is particularly relevant to the forthcoming 5G HetNet scenario where

MTs are expected to have very high data rate request.

In the considered simulation environment, both MeNBs and FAPs are assumed to use proportional fairness (PF) [111] as their MAC access mechanism. According to PF, RB k of eNB x is allocated to MT y^* if the metric m_{xy}^k attains the maximum value for $y = y^*$. Here m_{xy}^k can be computed as: $m_{xy}^k = \frac{d_{xy}^k(t)}{R_{xy}(t-1)}$ where $d_{xy}^k(t)$ denotes the achievable data rate by MT y from eNB x over RB k at time instant t and $R_{xy}(t-1)$ denotes the past average data rate achieved by MT y from eNB x upto time instant $t-1$. Here $d_{xy}^k(t)$ has been estimated from received SINR values. To deal with the ping-pong effect, we have used standard moving average method [114] to compute $R_{xy}(t-1)$ instead of relying on instantaneous data rate. The SINR received by MT y from eNB x through sub-carrier c has been computed as [85]:

$$\gamma_{xy}^c = \frac{P_{xy}^c}{FN_0B + I_y^c}. \tag{6.17}$$

Here P_{xy}^c denotes the received power by MT y from eNB x through sub-carrier c, F denotes the noise figure, N_0 denotes the noise spectral density, B denotes the bandwidth of an RB and I_y^c denotes the interference perceived by MT y through sub-carrier c.

To compute the effective SINR perceived by an MT through the sub-carriers of RB k, we have employed the exponential effective SINR mapping (EESM) model which is frequently used as a physical layer abstraction tool in system level simulations [115], [116]. Using EESM model, the effective SINR perceived at an MT through the sub-carriers of RB k can be computed as:

$$\gamma_{xy}^{eff} = -\beta \ln\left(\frac{1}{N_k}\sum_{c=1}^{N_k} exp\left(-\frac{\gamma_{xy}^c}{\beta}\right)\right).$$

Here N_k represents the total number of sub-carriers in RB k and β is an SINR scaling parameter depending on the modulation and coding scheme (MCS). The values of β for different MCS can be found in [116]. It may be noted that in EESM model, same MCS is used for all the sub-carriers allocated to a particular MT. Finally, cumulative throughput achieved by MT y from eNB x through RB k has been computed as [59]:

$$D_{xy}^k = \frac{N_k S_{eff}}{T_s}\log_2\left(1+\gamma_{xy}^{eff}\right).$$

Table 6.2: Parameter settings to evaluate p

Parameter	Value	Parameter	Value
F [85]	2.5	N_0 [85]	-174 dBm
B [85]	180 KHz	N_k [59]	12
S_{eff} [59]	10 symbols	T_s [59]	1 ms
α_{on} [51]	1.1	α_{off} [51]	1.5
γ [51]	1.0	λ [51]	$6 \sim 16$
v	50 Km/h	R	100 meter

Here N_k is the number of sub-carriers in RB k, S_{eff} is the number effective OFDMA symbols per LTE−A frame in time domain and T_s is the frame length. It may be noted that FAPs belonging to radio access technologies other than LTE−A (e.g. $802.11n$ WLAN) can be easily added in our simulation environment by configuring the base stations appropriately. The physical rate obtained from $802.11n$ FAPs can be derived from the received signal strength indicator (RSSI) values as described in [80]. Further, actual throughput obtained from a FAP can be measured as the ratio of the physical data rate to the total number of MTs associated with that FAP. Important parameter values used in our simulations are depicted in Table 6.2. The source code of our developed simulator is available at [117]. Based on this simulation framework, in the next section, the values of p for different traffic load have been computed.

6.2.1.1 Evaluating p for different traffic load conditions

Figure 6.7 depicts the effect of network traffic load on handover failure probability. Here the network traffic load varies from 500 MTs to 1300 MTs with a step of 100 MTs. The parameter of the Poisson distribution modeling CBR traffic arrival rate has been set to 7. The result shows that the outage probability in semisoft handover is considerably lower (performance gain 42%-46% approx.) compared to that of the hard handover. This performance gain for semisoft handover shows a decreasing trend with increasing traffic load. This is because the L2 switching delay associated with hard handover is considerably higher compared to that of semisoft handover. Since hard is a break-before-make mechanism, the MT cannot receive any data from the system during handover. This leads to severe degradation of user perceived throughput. In contrast, in semisoft handover, the

MT can concurrently maintain control links with both the current and target eNBs which minimizes handover latency as well as user throughput degradation during handover. Consequently, throughput degradation in semisoft handover is much lower compared to that of hard handover. As the traffic load becomes higher, achievable throughput from both the current and target eNBs degrades due to co-channel interference perceived from neighboring cells. Apart from that forwarding all user plane packets from current eNB to target eNB causes high congestion delay and packet losses under extreme load. As a result, semisoft handover cannot arrest the downfall of user perceived throughput when the traffic load is very high. The simulation results presented here have been obtained through Monte Carlo simulation method. The 99% confidence interval (CI) of the simulation results originate from $10,000$ independent runs of the simulator. CIs along with margin of errors (MOE) in measurements of p under different traffic load conditions have been depicted in Table 6.4.

6.2.2 Analysing the combined effect based on the analytical models

In this section, handover performances of different combinations of MMPs and HEMs have been evaluated based on the performance evaluation metrics presented in subsections 6.1.2, 6.1.3 and 6.1.4. In this analysis, we have used the p values obtained from simulation as input. We have used standard parameter values [19], [21], [118] to analyze handover latencies of FMIPv6 and SIGMA. The considered parameter values are depicted in Table 6.3. Figure 6.8 shows the effect of various HEMs on handover latency

Table 6.3: Parameter settings to evaluate handover latency

Parameter	Value
T_{L2}	45.35 ms
E_s	10 packets
T_{DAD}, T_{conf}	1000 ms
Bandwidth of wired links	100 Mbps
Processing delay at each hop	1 ms
Propagation delay of wired links	0.5 ms
Propagation delay of wireless links	10 ms
Encapsulation-decapsulation delay	1 ms

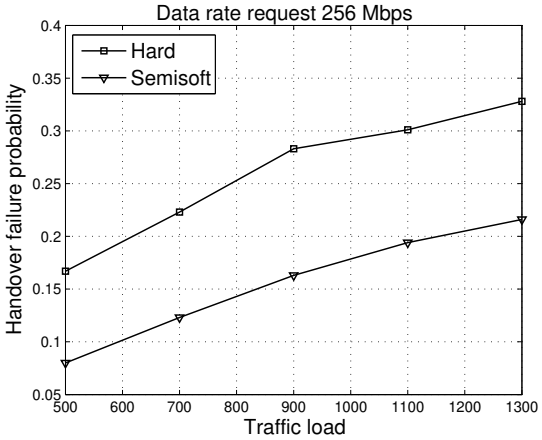


Figure 6.7: Handover failure probability vs. Traffic load

Table 6.4: CI and MOE measurements for p

Traffic load	Hard		Semisoft	
	CI (99%)	MOE	CI (99%)	MOE
500	(0.129, 0.205)	0.038	(0.037, 0.123)	0.0432
700	(0.19, 0.256)	0.032	(0.085, 0.161)	0.0383
900	(0.255, 0.311)	0.027	(0.130, 0.197)	0.0334
1100	(0.276, 0.325)	0.026	(0.165, 0.224)	0.0293
1300	(0.306, 0.349)	0.021	(0.190, 0.242)	0.0259

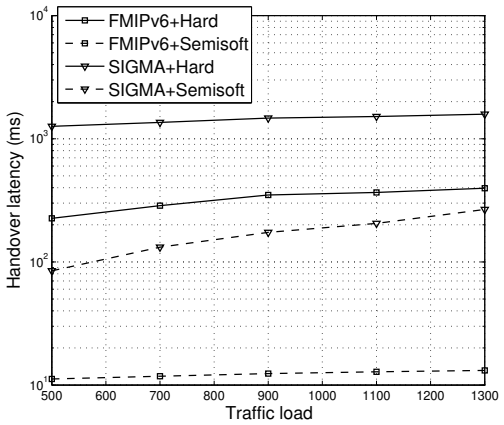


Figure 6.8: Handover latency (`FMIPv6` vs. `SIGMA`)

of different MMPs. The call arrival rate for CBR traffic varies from 7 to 11 units with a step of 1 unit while traffic load is set to 1000 MTs. The result shows that for a given handover execution mechanism, handover latency of FMIPv6 is significantly lower compared to that of SIGMA. It may be observed that the performance gain of FMIPv6 + hard over SIGMA + hard is approximately 82 % to 84 %, whereas the performance gain of FMIPv6 + Semisoft over SIGMA + Semisoft is approximately 86 % to 90 %. The reasons behind are as follows. In predictive mode operation of FMIPv6, delay associated with IP address configuration such as T_{DAD} is not included as a part of handover latency in FMIPv6 + hard. However, when handover failure occurs, the FMIPv6 protocol switches to reactive mode and T_{DAD} becomes a part of handover latency. In contrast to FMIPv6 + hard, the IP address configuration delay such as T_{conf} is an *essential* part of handover latency for SIGMA + hard. As a result, FMIPv6 + hard outperforms SIGMA + hard in terms of handover latency. On the other hand, in FMIPv6 + Semisoft, delay associated with IP address configuration such as T_{DAD} is not a part of handover latency at all. However, in SIGMA + Semisoft, T_{conf} becomes a part of handover latency in case of handover failure. As a result, FMIPv6 + Semisoft outperforms SIGMA + Semisoft. The result also shows that for a given MMP, semisoft handover significantly outperforms hard handover in terms of handover latency. It can be observed that SIGMA + Semisoft can achieve a performance gain of 95 % to 98 % compared to that of SIGMA + hard. On the other hand, FMIPv6 + Semisoft can achieve a performance gain of 92 % to 95 % compared to that of FMIPv6 + hard. This is because in FMIPv6 + Semisoft, T_{DAD} delay is completely eliminated whereas it becomes a part of handover latency in FMIPv6 + hard. Similarly, T_{update} and T_{conf} are included in handover latency when SIGMA + hard are used, whereas they are eliminated from handover latency when SIGMA + Semisoft is used. The result also shows that the superiority of FMIPv6 over SIGMA is actually conditional on the underlying handover execution mechanism. According to the analysis, FMIPv6 + hard outperforms SIGMA + hard, however SIGMA + Semisoft outperforms FMIPv6 + hard (performance gain 10 % to 30 % approximately). On the other hand, SIGMA+Semisoft outperforms FMIPv6+Hard, whereas FMIPv6+Semisoft outperforms SIGMA+Semisoft (performance gain 11 % to 20 % approximately). The reasons behind are similar to that described previously.

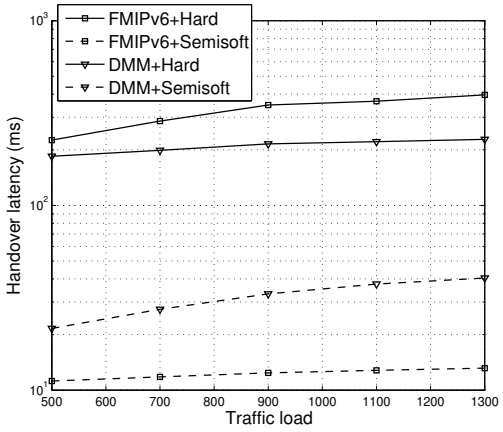


Figure 6.9: Handover latency (FMIPv6 vs. DMM)

This performance analysis reveals that the superiority of SIGMA over FMIPv6 is actually conditional on the underlying HEMs.

Figure 6.9 shows the effect of increasing traffic load on handover latency. Results show that when hard handover is used at L2, DMM protocol outperforms FMIPv6 in terms of handover latency (performance gain 15%-25% approximately). This is because in DMM protocol, delay due to DAD procedure is not part of handover latency when the handover is performed within same PMIPv6 domain. In contrast, DAD delay becomes part of handover latency of FMIPv6 in case of handover failure. It may also be observed that the performance gain of DMM over FMIPv6 increases with increasing traffic load when hard handover is used at L2. This is because the handover failure increases with increasing traffic load. When semi-soft handover is used at L2, then handover latency for both FMIPv6 and DMM decreases. This is because DAD latency is eliminated from the handover latency. In contrast to hard handover, DMM outperforms FMIPv6 when semisoft handover is used at L2. This is because of the increased delay in DMM due to extended PBU and PBA signaling. *This performance analysis reveals that the superiority of FMIPv6 over DMM is actually conditional on underlying HEMs.*

Figure 6.10 shows the effect of increasing traffic load on handover latencies of SIGMA and DMM. Results show that when hard handover in used at L2, DMM significantly outperforms SIGMA (performance gain upto 90% approximately). This is because SIGMA cannot exploit multihoming and multistreaming capacity of SCTP when

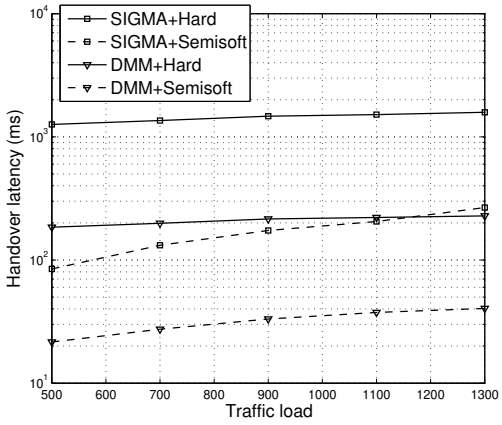


Figure 6.10: Handover latency (SIGMA vs. DMM)

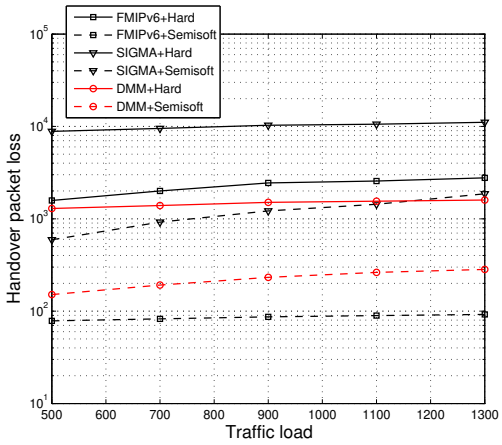


Figure 6.11: Handover packet loss vs. traffic load

hard handover is used at L2. As a result, IP address configuration delay becomes part of handover latency in SIGMA. In contrast, the handover latency of DMM protocol do not include such delay when the handover is performed within a PMIPv6 domain. As a result, the handover latency of DMM+Hard is much lower than SIGMA+Hard. When semisoft handover is used at L2, then the handover latency becomes much lower for both DMM and SIGMA compared to that when hard is used at L2. Here also DMM+Semisoft outperforms SIGMA+Semisoft. This is because in case of handover failure, the handover process of SIGMA includes IP address configuration delay whereas the DAD latency is completely eliminated from DMM. Apart from that the path switching delay in SIGMA

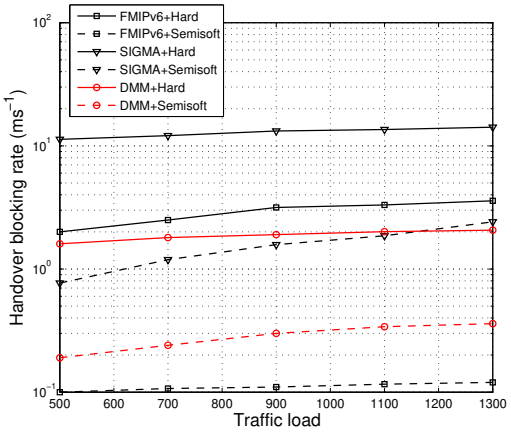


Figure 6.12: Handover blocking rate vs. traffic load

is higher compared to that of DMM. As a result, for a given handover execution mechanism, DMM outperforms SIGMA in terms of handover latency. The result also shows that the superiority of DMM over SIGMA is conditional on underlying handover execution mechanism for lower traffic load ($<$ 1150 MTs). Here DMM+Hard outperforms SIGMA+Hard, however SIGMA+Semisoft outperforms DMM+Hard. On the other hand, DMM always outperforms SIGMA as the traffic load exceeds 1150 MTs. This is due to rapid increase of handover failure under high traffic load condition. *This performance analysis reveals that the superiority of DMM over SIGMA is conditional on both traffic load and underlying HEMs.*

Figures 6.11 and 6.12 depict the effect of increasing traffic load on handover packet loss and handover blocking rate respectively. Results show that for a given MMP, handover packet loss is significantly reduced when semisoft handover is used at L2. This is due to the reduced handover latency incurred by semisoft handover mechanism compared to that of hard handover. Results also show that for a given handover execution mechanism, both DMM and FMIPv6 outperforms SIGMA. Moreover, superiority of FMIPv6 over DMM is conditional on underlying handover execution mechanism as long as traffic load is less than 1150 MTs. DMM always outperforms SIGMA as the traffic load increases beyond 1150 MTs. The reasons behind such phenomenon are similar to that described previously.

6.3 Proposed ranking scheme

Based on the combined effect presented in section 6.2, in this section we prioritize between different combinations of MMPs and HEMs employing analytic hierarchy process (AHP). AHP is a well known mathematical tool to select the best option from a *set of alternative options* based on a *set of evaluation criteria* [119], [120]. The goal of AHP is to find the most suitable trade-off among the considered evaluation criteria. In our case, the set of alternative options consists of six different combinations namely FMIPv6+Hard, FMIPv6+Semisoft, DMM+Hard, DMM+Semisoft, SIGMA+Hard and SIGMA+Semisoft. On the other hand, the set of alternative evaluation criteria consists of three different metrics namely handover latency, handover packet loss and handover blocking rate. While evaluating different evaluation criteria, we consider the requirements for video streaming applications. Streaming applications need high data rate and reliability, however they are not affected by delay [119]. Different steps of the adopted AHP are described below:

6.3.1 Computing the vector of criteria weights

In this phase, we construct the pairwise comparison matrix **A** based on pairwise comparisons of the considered criteria, i.e., handover latency, handover packet loss and handover blocking rate. Here **A** is a 3×3 real matrix where each entry a_{jk} represents the relative importance of j th criterion with respect to the k th criterion $(1 \leq j \leq 3, 1 \leq k \leq 3)$. In this matrix, $a_{jk} > 1$ means criterion j is *more* important than criterion k. On the other hand, $a_{jk} < 1$ means criterion j is *less* important than criterion k. Here $a_{jk} = 1$ when j is *equally* important as k. Relative importances between pairs of criteria have been measured according to the well known numerical scale from 1 to 9 [119], [120]. Since streaming applications have strict data rate requirements, handover blocking rate has been assigned highest priority. Further, handover packet loss has been assigned higher priority compared to handover latency. This is because streaming applications require high reliability whereas they are not severely affected by delay. The matrix **A** considered in our analysis is depicted in Table 6.5. To check the consistency of the pairwise evaluations made in **A**, we have computed consistency ratio (CR) which is a well known metric in AHP. The CR of **A** $(= 0.043)$ reveals that the pairwise evaluations are consistent as

Table 6.5: Pairwise comparison matrix **A** and weight vector **W** (**CR** $= 0.043 < 0.1$)

	Blocking rate	Packet loss	Latency	**W**
Blocking rate	1	3	5	1.90
Packet loss	$\frac{1}{3}$	1	3	0.77
Latency	$\frac{1}{5}$	$\frac{1}{3}$	1	0.31

the value of CR is less than 0.1. From the matrix **A**, the normalized pairwise comparison matrix **N** is derived by making the sum of entries on each column equal to 1. Clearly, **N** is also a 3×3 real matrix where each element n_{jk} $(1 \leq j \leq 3, 1 \leq k \leq 3)$ is computed as $n_{jk} = \frac{a_{jk}}{\sum_{i=1}^{3} a_{ik}}$. Further, we compute the criteria weight vector **W** by averaging the entries on each row of **N**. Here **W** is a 3×1 matrix where each element $w_j = \frac{1}{3}\sum_{k=1}^{3} n_{jk}$ (depicted in Table 6.5).

6.3.2 Computing the matrix of option scores

In this phase, for a fixed criterion, a score to each option has been assigned based on the pairwise comparisons of the options with respect to the considered criterion. Pairwise comparison matrices for handover blocking rate, handover packet loss and handover latency are depicted in Tables 6.6, 6.7 and 6.8 respectively. These matrices have been constructed similar to that of **A** as described in section 6.3.1. Since we have six options, each pairwise comparison matrix is a 6×6 matrix. *Here the relative importances of one option with respect to the others have been determined based on the results obtained from our proposed analytical framework (presented in section 6.2.2).* While determining the weights, we consider the performances of the considered options under high traffic load condition (> 1150 MTs) which is more relevant to the UDN scenario. The CR for each of these matrices establish the consistency of the pairwise evaluations. From these pairwise comparison matrices, we have computed the score vectors **X**, **Y** and **Z** following the similar method of computing **W**. Here the vectors **X**, **Y** and **Z** represent the scores for different options when handover blocking rate, handover packet loss and handover latency are of primary concerns respectively. Finally, the score matrix **S** has

been computed as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} & \mathbf{Z} \end{bmatrix} \tag{6.18}$$

i.e., the first, second and third columns of **S** corresponds to the vectors **X**, **Y** and **Z** respectively. Clearly, **S** is a 6×3 matrix.

6.3.3 Ranking the options

In this phase, we obtain the vector **V** of global scores by multiplying **S** and **W**, i.e., $\mathbf{V} = \mathbf{S} \bullet \mathbf{W}$. Here the i th entry of **V** represents the global score assigned by the AHP to the i th option. Finally, we obtain the option ranks by sorting the options based on global scores in decreasing order. The option ranks have been depicted in Table 6.9. Such ranking reveals the strong impact of underlying handover execution mechanism on FMIPv6 protocol in UDN scenario. It can be observed that the FMIPv6 protocol when used with semisoft handover mechanism significantly outperforms all other considered MMPs. The obtained ranking also shows that the DMM protocol always outperforms SIGMA irrespective of underlying HEMs. Furthermore, the rank of SIGMA+Semisoft is better than FMIPv6+Hard but their ranks are very close. From this analysis we argue that the network layer MMPs such as FMIPv6 and DMM generally outperforms transport layer MMPs such as SIGMA in UDN scenario. Such analysis of MMPs can be easily extended for any kind of applications (e.g., conversational and best effort services) and HEMs by constructing the comparison matrices consistently.

Table 6.6: Pairwise comparison matrix for handover blocking rate
(CR $= 0.0046 < 0.1$)

Criterion: Handover blocking rate	FMIPv6 + Hard	FMIPv6 + Semisoft	SIGMA + Hard	SIGMA + Semisoft	DMM + Hard	DMM + Semisoft	**X**
FMIPv6+Hard	1	0.11	2	0.83	0.66	0.4	0.41
FMIPv6+Semisoft	9	1	11	7	6	4	3.38
SIGMA+Hard	0.5	0.09	1	0.5	0.4	0.35	0.25
SIGMA+Semisoft	1.2	0.14	2	1	0.9	0.7	0.51
DMM+Hard	1.5	0.16	2.5	1.11	1	0.8	0.61
DMM+Semisoft	2.5	0.25	2.85	1.42	1.25	1	0.81

Table 6.7: Pairwise comparison matrix for handover packet loss
(CR $= 0.078 < 0.1$)

Criterion: Handover packet loss	FMIPv6 + Hard	FMIPv6 + Semisoft	SIGMA + Hard	SIGMA + Semisoft	DMM + Hard	DMM + Semisoft	**Y**
FMIPv6+Hard	1	0.55	1.2	0.625	0.5	0.25	0.497
FMIPv6+Semisoft	8	1	10	5	4	2	2.694
SIGMA+Hard	0.83	0.1	1	0.5	0.4	0.2	0.271
SIGMA+Semisoft	1.6	0.2	2	1	0.8	0.4	0.538
DMM+Hard	2	0.25	2.5	1.25	1	0.5	0.673
DMM+Semisoft	4	0.5	5	2.25	2	1	1.323

Table 6.8: Pairwise comparison matrix for handover latency
(CR $= 0.024 < 0.1$)

Criterion: Handover packet loss	FMIPv6 + Hard	FMIPv6 + Semisoft	SIGMA + Hard	SIGMA + Semisoft	DMM + Hard	DMM + Semisoft	**Z**
FMIPv6+Hard	1	0.125	2	1.5	0.5	0.33	0.45
FMIPv6+Semisoft	8	1	11	6	7	3	3.175
SIGMA+Hard	0.5	0.09	1	0.5	0.66	0.25	0.266
SIGMA+Semisoft	0.66	0.16	2	1	1.5	0.5	0.514
DMM+Hard	2	0.14	1.5	0.66	1	0.4	0.483
DMM+Semisoft	3	0.33	4	2	2.5	1	1.106

Table 6.9: Priority of different options in decreasing order of scores

Options	Score
FMIPv6+Semisoft	9.476
DMM+Semisoft	2.896
DMM+Hard	1.804
SIGMA+Semisoft	1.535
FMIPv6+Hard	1.295
SIGMA+Hard	0.7635

6.4 Conclusions

In this chapter, we have analyzed the handover performances of a network layer (FMIPv6), a transport layer (SIGMA) and a distributed mobility management (DMM) protocol for a dense deployment scenario of LTE-A HetNet. In this analysis, we explicitly consider the effect of underlying handover execution mechanisms such as hard and semisoft. From the results obtained from analyses, we conclude the following:

- Superiority of semisoft over hard is *conditional* on network traffic load in LTE-A HetNet. For lightly loaded situation, semisoft handover significantly outperforms hard handover. As the traffic load increases, the performance gain of semisoft over hard shows a decreasing trend.
- Superiority of SIGMA over FMIPv6 is actually conditional on the underlying handover execution mechanisms under dense deployment scenario.
- The performance of DMM is better compared to that of SIGMA irrespective of underlying handover execution mechanism.
- The superiority of DMM over FMIPv6 is conditional on underlying handover execution mechanism.

Since IP address configuration delay is the principle part of handover latency and the inclusion of such delay is determined by underlying handover execution mechanism, we argue that similar observation hold for other mobility management protocols operating from network layer, transport layer and exhibiting distributed mobility architecture.

Chapter 7

Conclusions and Future Directions

In this thesis, some analytical frameworks have been developed to analyze the effect of different system parameters on handover performances in heterogeneous network and based on such frameworks, some efficient handover algorithms have been proposed. Our study starts with an analytical framework to investigate the effect of medium access control mechanisms, upper layer mobility management protocols and link layer handover decision metrics on user perceived throughput. This analysis reveals that handover decision metric plays a crucial role in determining user perceived throughput in *all-IP* HetNet. Subsequently, we develop two handover decision metrics to deal with the blackout period in UDN scenario and handover anomaly problem in unlicensed band communications respectively. We also develop an analytical framework to investigate the performances of DC in C/U split HetNet. Our analyses reveal the conditional effect of traffic load density and call arrival rates on the performance gain of DC. Finally, we investigate the interdependence between handover execution mechanisms and upper layer MMPs. Based on such analysis, different combinations of MMPs and handover execution mechanisms have been prioritized using analytic hierarchy process. The significance of the developed analytical frameworks lies in the fact that these analyses would provide deeper insight towards service guarantee and system design in HetNet scenario.

In future, we aim to develop energy efficient component carrier selection algorithms for the newly standardized DC mechanism between LTE-A and new radio (NR). Since, the transmitting powers of macrocells are much higher compared to that of small cells, the MTs often can not utilize the offloading capacity of small cell NR base stations. This

causes load imbalance across the serving cells. Such load imbalance may cause high queuing delay at the serving macrocell. Moreover, to minimize energy consumption in DC mode, component carrier allocation per user need to be minimized. Towards this end, some new component carrier selection algorithms for LTE−NR heterogeneous networks will be developed. Our goal is to minimize the number of component carrier allocations considering load balancing across the serving cells.

Bibliography

[1] Y. Li, B. Cao, and C. Wang. Handover schemes in heterogeneous lte networks: challenges and opportunities. *IEEE Wireless Communications*, 23(2):112–117, 2016.

[2] A. Gupta and R. K. Jha. A survey of 5g network: Architecture and emerging technologies. *IEEE Access*, 3:1206–1232, 2015.

[3] M. Kamel, W. Hamouda, and A. Youssef. Ultra-dense networks: A survey. *IEEE Communications Surveys Tutorials*, 18(4):2522–2545, 2016.

[4] L. Du, N. Zheng, H. Zhou, J. Chen, T. Yu, X. Liu, Y. Liu, Z. Zhao, X. Qian, J. Chi, Z. Chen, and G. Liu. C/u split multi-connectivity in the next generation new radio system. In *IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2017.

[5] H. Wang, C. Rosa, and K. I. Pedersen. Dual connectivity for lte-advanced heterogeneous networks. *Wireless Networks*, 22(4):1315–1328, 2016.

[6] Y. Khadraoui, X. Lagrange, and A. Gravey. Very tight coupling between lte and wifi: From theory to practice. In *Wireless Days (WD)*, pages 1–3, 2016.

[7] X. Lagrange. Very tight coupling between lte and wi-fi for advanced offloading procedures. In *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 82–86, 2014.

[8] S. Ferretti, V. Ghini, and F. Panzieri. A survey on handover management in mobility architectures. *Computer Networks*, 94(C):390–413, 2016.

[9] R. B. Ali and S. Pierre. On the impact of soft vertical handoff on optimal voice admission control in pcf-based wlans loosely coupled to 3g networks. *IEEE Transactions on Wireless Communications*, 8(3):1356–1365, 2009.

[10] D. Saha, A. Mukherjee, I. S. Misra, and M. Chakraborty. Mobility support in IP: a survey of related protocols. *IEEE Network*, 18(6):34–40, 2004.

[11] R. Ahmad, E. A. Sundararajan, N. E. Othman, and M. Ismail. Handover in lte-advanced wireless networks: state of art and survey of decision algorithm. *Telecommunication Systems*, 66(3):533–558, 2017.

[12] Technical Specification Group Radio Access Network 3GPP. Version 6.7.0. In *TS 25.214. Physical Layer Procedures (FDD)*, 2006.

[13] S. C. Ghosh, R. M. Whitaker, S. M. Allen, and S. Hurley. Optimising cdma cell planning with soft handover. *Wireless Personal Communication*, 68(3):321–347, 2013.

[14] D. G. Brennan. Linear diversity combining techniques. *Proceedings of the IEEE*, 91(2):331–356, 2003.

[15] H. Lee, H. Son, and S. Lee. Semisoft handover gain analysis over ofdm-based broadband systems. *IEEE Transactions on Vehicular Technology*, 58(3):1443–1453, 2009.

[16] J. Chang, Y. Li, S. Feng, H. Wang, C. Sun, and P. Zhang. A fractional soft handover scheme for 3gpp lte-advanced system. In *IEEE International Conference on Communications Workshops*, pages 1–5, 2009.

[17] R. P. Antonioli et al. Dual connectivity for lte-nr cellular networks. In *XXXV SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2017*, 2017.

[18] C. Makaya and S. Pierre. An analytical framework for performance evaluation of ipv6-based mobility management protocols. *IEEE Transactions on Wireless Communications*, 7(3):972–983, 2008.

[19] J. Lee, J. Bonnin, I. You, and T. Chung. Comparative handover performance analysis of ipv6 mobility management protocols. *IEEE Transactions on Industrial Electronics*, 60(3):1077–1088, 2013.

[20] K. Kong, W. Lee, Y. Han, M. Shin, and H. You. Mobility management for all-ip mobile networks: mobile ipv6 vs. proxy mobile ipv6. *IEEE Wireless Communications*, 15(2):36–45, 2008.

[21] J. Lee, T. Ernst, and T. Chung. Cost analysis of ip mobility management protocols for consumer mobile devices. *IEEE Transactions on Consumer Electronics*, 56(2):1010–1017, 2010.

[22] F. Giust, L. Cominardi, and C. J. Bernardos. Distributed mobility management for future 5g networks: overview and analysis of existing approaches. *IEEE Communications Magazine*, 53(1):142–149, 2015.

[23] Md. S. Hossain and M. Atiquzzaman. Cost analysis of mobility protocols. *Telecommunication Systems*, 52(4):2271–2285, 2013.

[24] A. Mahmoud, A. Al-Helali, M. Abu-Amara, T. Al-Kharobi, and T. Sheltami. Comparative performance study for integrated 3g/wlan networks using mobile ip, sip, and m-sctp protocols. In *IEEE 71st Vehicular Technology Conference*, pages 1–5, 2010.

[25] Shaojian Fu and Mohammed Atiquzzaman. Survivability evaluation of SIGMA and mobile IP. *Wireless Personal Communication*, 43(3):933–944, 2007.

[26] S. Mohanty and I. F. Akyildiz. Performance analysis of handoff techniques based on mobile ip, tcp-migrate, and sip. *IEEE Transactions on Mobile Computing*, 6(7):731–747, 2007.

[27] B. Sardar and D. Saha. Performance analysis of basic support protocol (bsp) in nested network mobility (nenemo). In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1430–1435, 2013.

[28] A. Kumar and V. Kumar. Optimal association of stations and aps in an ieee 802.11 wlan. In *National Communications Conference (NCC)*, 2005.

[29] S. C. Ghosh, R. M. Whitaker, S. M. Allen, and S. Hurley. Dynamic data resolution to improve the tractability of UMTS network planning. *Annals of Operations Research*, 201(1):197–227, 2012.

[30] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda. Downlink packet scheduling in lte cellular networks: Key design issues and a survey. *IEEE Communications Surveys Tutorials*, 15(2):678–700, 2013.

[31] P. K. Chowdhury, A. S. Reaz, M. Atiquzzaman, and W. Ivancic. Performance analysis of sinemo: Seamless ip-diversity based network mobility. In *IEEE International Conference on Communications (ICC)*, pages 6032–6037, 2007.

[32] M. Laddomada, F. Mesiti, M. Mondin, and F. Daneshgaran. On the throughput performance of multirate ieee 802.11 networks with variable-loaded stations: analysis, modeling, and a novel proportional fairness criterion. *IEEE Transactions on Wireless Communications*, 9(5):1594–1607, 2010.

[33] H. Wang, D. I. Laurenson, and J. Hillston. A general performance evaluation framework for network selection strategies in 3g-wlan interworking networks. *IEEE Transactions on Mobile Computing*, 12(5):868–884, 2013.

[34] R. Ahmad, E. A. Sundararajan, and A. F. Khalifeh. A survey on femtocell handover management in dense heterogeneous 5g networks. *Telecommunication Systems*, 75(4):481–507, 2020.

[35] A. Orsino, G. Araniti, A. Molinaro, and A. Iera. Effective rat selection approach for 5g dense wireless networks. In *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2015.

[36] S. Boumard, I. Harjula, K. Horneman, and H. Hu. Throughput and energy consumption trade-off in traffic splitting in heterogeneous networks with dual connectivity. In *28th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–5, 2017.

[37] P. Hsieh, W. Lin, K. Lin, and H. Wei. Dual-connectivity prevenient handover scheme in control/user-plane split networks. *IEEE Transactions on Vehicular Technology*, 67(4):3545–3560, 2018.

[38] C. Cano, D. Lopez-Perez, H. Claussen, and D. J. Leith. Using lte in unlicensed bands: Potential benefits and coexistence issues. *IEEE Communications Magazine*, 54(12):116–123, 2016.

[39] Z. Guan and T. Melodia. Cu-lte: Spectrally-efficient and fair coexistence between lte and wi-fi in unlicensed bands. In *35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, pages 1–9, 2016.

[40] A. Habbal, S. I. Goudar, and S. Hassan. Context-aware radio access technology selection in 5g ultra dense networks. *IEEE Access*, 5:6636–6648, 2017.

[41] F. Bouali, K. Moessner, and M. Fitch. A context-aware user-driven framework for network selection in 5g multi-rat environments. In *IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–7, 2016.

[42] A. Kaloxylos, S. Barmpounakis, P. Spapis, and N. Alonistioti. An efficient rat selection mechanism for 5g cellular networks. In *International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 942–947, 2014.

[43] X. Wang, M. Xiao, J. Yi, C. Feng, and F. Jiang. On the performance analysis of downlink heterogeneous networks with dual connectivity. In *8th International Conference on Wireless Communications Signal Processing (WCSP)*, pages 1–6, 2016.

[44] L. Yan, X. Fang, and Y. Fang. A novel network architecture for c/u-plane staggered handover in 5g decoupled heterogeneous railway wireless systems. *IEEE Transactions on Intelligent Transportation Systems*, 18(12):3350–3362, 2017.

[45] J. Zhao, Y. Liu, Y. Gong, C. Wang, and L. Fan. A dual-link soft handover scheme for c/u plane split network in high-speed railway. *IEEE Access*, 6:12473–12482, 2018.

[46] C. J. Bernardos, A. Oliva, and F. Giust. A pmipv6-based solution for distributed mobility management. In *IETF Draft, draft-bernardos-dmm-pmip-09*, 2017.

[47] S. K. Ghosh and S. C. Ghosh. An analytical framework for throughput analysis of real time applications in all-ip networks. In *IEEE 31st International Conference*

on Advanced Information Networking and Applications (AINA), pages 508–515, 2017.

[48] L. Ma, F. R. Yu, and V. C. M. Leung. Performance improvements of mobile sctp in integrated heterogeneous wireless networks. *IEEE Transactions on Wireless Communications*, 6(10):3567–3577, 2007.

[49] X. Wang, T. Q. S. Quek, M. Sheng, and J. Li. Throughput and fairness analysis of wi-fi and lte-u in unlicensed band. *IEEE Journal on Selected Areas in Communications*, 35(1):63–78, 2017.

[50] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. Wong, R. Schober, and L. Hanzo. User association in 5g networks: A survey and an outlook. *IEEE Communications Surveys Tutorials*, 18(2):1018–1044, 2016.

[51] B. Chang and J. Chen. Cross-layer-based adaptive vertical handoff with predictive rss in heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, 57(6):3679–3692, 2008.

[52] J. Chen, Z. Wei, Y. Wang, L. Sang, and D. Yang. A service-adaptive multi-criteria vertical handoff algorithm in heterogeneous wireless networks. In *IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 899–904, 2012.

[53] H. Zhou, S. Mao, and P. Agrawal. Approximation algorithms for cell association and scheduling in femtocell networks. *IEEE Transactions on Emerging Topics in Computing*, 3(3):432–443, 2015.

[54] S. K. Ghosh and S. C. Ghosh. A goodness based vertical handoff algorithm for heterogeneous networks. In *14th IFIP Wired/Wireless Internet Communications*, pages 254–267, 2016.

[55] A. Nazari, P. Branch, J. But, and H. L. Vu. Conservative soft handoff for heterogeneous wireless networks. In *IEEE Wireless Communication and Networking Conference*, pages 1–6, 2010.

[56] C. Rosa, K. I. Pedersen, H. Wang, P. Michaelsen, S. Barbera, E. Malkamäki, T. Henttonen, and B. Sébire. Dual connectivity for LTE small cell evolution: functionality and performance aspects. *IEEE Communications Magazine*, 54(6):137–143, 2016.

[57] S. C. Jha, K. Sivanesan, R. Vannithamby, and A. Taha Koç. Dual connectivity in lte small cell networks. In *2014 IEEE Global Communications Conference (GLOBECOM) Workshops*, pages 1205–1210, 2014.

[58] N. Prasad and S. Rangarajan. Exploiting dual connectivity in heterogeneous cellular networks. In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 1–8, 2017.

[59] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi. Improved handover through dual connectivity in 5g mmwave mobile networks. *IEEE Journal on Selected Areas in Communications*, 35(9):2069–2084, 2017.

[60] S. Altamimi and S. Shirmohammadi. Qoe-fair dash video streaming using server-side reinforcement learning. *ACM Transactions on Multimedia Computing, Communications and Applications*, 16(2s), 2020.

[61] B. Wang, J. Kurose, P. Shenoy, and D. Towsley. Multimedia streaming via tcp: An analytic performance study. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 4(2), 2008.

[62] A. Martin, J. Egaña, J. Flórez, J. Montalbán, I. G. Olaizola, M. Quartulli, R. Viola, and M. Zorrilla. Network resource allocation system for qoe-aware delivery of media services in 5g networks. *IEEE Transactions on Broadcasting*, 64(2):561–574, 2018.

[63] A. Ahmed, L. M. Boulahia, and D. Gaïti. Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification. *IEEE Communications Surveys Tutorials*, 16(2):776–811, 2014.

[64] A. D. L. Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal. An overview of ieee 802.21: media-independent handover services. *IEEE Wireless Communications*, 15(4):96–103, 2008.

[65] M. Pan, T. Lin, C. Chiu, and C. Wang. Downlink traffic scheduling for LTE-A small cell networks with dual connectivity enhancement. *IEEE Communications Letters*, 20(4):796–799, 2016.

[66] A. Zakrzewska, D. López-Pérez, S. Kucera, and H. Claussen. Dual connectivity in lte hetnets with split control- and user-plane. In *IEEE Globecom Workshops (GC Wkshps)*, pages 391–396, 2013.

[67] G. Pocovi, S. Barcos, H. Wang, K. I. Pedersen, and C. Rosa. Analysis of heterogeneous networks with dual connectivity in a realistic urban deployment. In *IEEE 81st Vehicular Technology Conference (VTC)*, pages 1–5, 2015.

[68] H. He, X. Li, Z. Feng, J. Hao, X. Wang, and H. Zhang. An adaptive handover trigger strategy for 5g c/u plane split heterogeneous network. In *IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 476–480, 2017.

[69] P. Kundu, S. K. Ghosh, and B. Sardar. Performance analysis of sinemo and nemo bsp for vehicle-infrastructure-vehicle communications. In *Applications and Innovations in Mobile Computing (AIMoC)*, pages 75–80, 2014.

[70] S. K. Ghosh, P. Kundu, B. Sardar, and D. Saha. An extension of on-board tcp (obtcp) for satellite-terrestrial hybrid networks. In *4th International Conference of Emerging Applications of Information Technology*, pages 146–151, 2014.

[71] T. Do and Y. Kim. Design and performance analysis of an energy-efficient uplink carrier aggregation scheme. *Wireless Network*, 21(7):2303 – 2314, 2015.

[72] J. Carmona-Murillo, V. Friderikos, and J. L. G. Sánchez. A hybrid DMM solution and trade-off analysis for future wireless networks. *Computer Networks*, 133:17–32, 2018.

[73] T. M. Ali and M. Saquib. Analytical framework for wlan-cellular voice handover evaluation. *IEEE Transactions on Mobile Computing*, 12(3):447–460, 2013.

[74] N. K. Panigrahy and S. C. Ghosh. Analyzing the effect of soft handover on wlan usage efficiency under load condition. pages 192–199, 2015.

[75] A. H. Zahran, B. Liang, and A. Saleh. Mobility modeling and performance evaluation of heterogeneous wireless networks. *IEEE Transactions on Mobile Computing*, 7(8):1041–1056, 2008.

[76] K. Yang, I. Gondal, B. Qiu, and L. S. Dooley. Combined sinr based vertical handoff algorithm for next generation heterogeneous wireless networks. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 4483–4487, 2007.

[77] W. Stallings. *Data and Computer Communications*. Prentice Hall Press, Upper Saddle River, NJ, USA, 10th edition, 2013.

[78] S. K. Ghosh and S. C. Ghosh. A goodness based vertical handoff algorithm for heterogeneous networks. In *14th IFIP Wired/Wireless Internet Communications*, pages 254–267, 2016.

[79] C. Bettstetter. Smooth is better than sharp: A random mobility model for simulation of wireless networks. In *4th ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 19–27, 2001.

[80] D. Gong and Y. Yang. Ap association in 802.11n wlans with heterogeneous clients. In *IEEE IEEE Conference on Computer Communications (INFOCOM)*, pages 1440–1448, 2012.

[81] C. Lung and C. Zhou. Using hierarchical agglomerative clustering in wireless sensor networks: An energy-efficient and flexible approach. *Ad Hoc Networks*, 8(3):328 – 344, 2010.

[82] S. Stefanatos and A. Alexiou. Access point density and bandwidth partitioning in ultra dense wireless networks. *IEEE Transactions on Communications*, 62(9):3376–3384, 2014.

[83] S. K. Ghosh and S. C. Ghosh. Analyzing the performance of dual connectivity in control/user-plane split heterogeneous networks. In *15th IEEE/IFIP Wireless On-demand Network systems and Services Conference (WONS)*, pages 64–71, 2019.

[84] C. Bouras, G. Diles, V. Kokkinos, K. Kontodimas, and A. Papazois. A simulation framework for evaluating interference mitigation techniques in heterogeneous cellular environments. *Wireless Personal Communications*, 77(2):1213–1237, 2014.

[85] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda. Simulating lte cellular systems: An open-source framework. *IEEE Transactions on Vehicular Technology*, 60(2):498–513, 2011.

[86] ETSI TS 123 203. Lte: Policy and charging control architecture, 2012.

[87] 3GPP-TS 36.300 v8.5.0. E-utran overall description, 2008.

[88] A. Rao, A. Legout, Y. Lim, D. Towsley, C. Barakat, and W. Dabbous. Network characteristics of video streaming traffic. In *Proceedings of the Seventh Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, pages 25:1–25:12, New York, NY, USA, 2011. ACM.

[89] J. Zeng, O. C. Au, W. Dai, Y. Kong, L. Jia, and W. Zhu. A tutorial on image/video coding standards. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–7, 2013.

[90] X. Lagrange. Very tight coupling between lte and wi-fi for advanced offloading procedures. In *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 82–86, 2014.

[91] L. Wang, Y. Zhang, and Z. Wei. Mobility management schemes at radio network layer for lte femtocells. In *IEEE 69th Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2009.

[92] A. Prasad, O. Tirkkonen, P. Lundén, O. N. C. Yilmaz, L. Dalsgaard, and C. Wijting. Energy-efficient inter-frequency small cell discovery techniques for lte-advanced heterogeneous network deployments. *IEEE Communications Magazine*, 51(5):72–81, 2013.

[93] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of american statistical association*, 58(301), 1963.

[94] S. Sagari, I. Seskar, and D. Raychaudhuri. Modeling the coexistence of lte and wifi heterogeneous networks in dense deployment scenarios. In *IEEE International Conference on Communication Workshop (ICCW)*, pages 2301–2306, 2015.

[95] N. C. Sagias, G. S. Tombras, and G. K. Karagiannidis. New results for the shannon channel capacity in generalized fading channels. *IEEE Communications Letters*, 9(2):97–99, 2005.

[96] A. V. Babu and L. Jacob. Fairness analysis of ieee 802.11 multirate wireless lans. *IEEE Transactions on Vehicular Technology*, 56(5):3073–3088, 2007.

[97] H. Bateman. Higher transcendal functions, 2016.

[98] M. M. Jamei and W. Koepf. A new identity for generalized hypergeometric functions and applications. *Axioms*, 8(1), 2019.

[99] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli. Control-data separation architecture for cellular radio access networks: A survey and outlook. *IEEE Communications Surveys Tutorials*, 18(1):446–465, 2016.

[100] H. Ishii, Y. Kishiyama, and H. Takahashi. A novel architecture for LTE-B : C-plane/u-plane split and phantom cell concept. In *Workshops Proceedings of the Global Communications Conference (GLOBECOM)*, pages 624–630, 2012.

[101] L. Irio, R. Oliveira, and L. Bernardo. Aggregate interference in random waypoint mobile networks. *IEEE Communications Letters*, 19(6):1021–1024, 2015.

[102] J. G. Andrews, F. Baccelli, and R. K. Ganti. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11):3122–3134, 2011.

[103] S. Agarwal, S. De, S. Kumar, and H. M. Gupta. Qos-aware downlink cooperation for cell-edge and handoff users. *IEEE Transactions on Vehicular Technology*, 64(6):2512–2527, 2015.

[104] H. Fu and D. I. Kim. Analysis of throughput and fairness with downlink scheduling in wcdma networks. *IEEE Transactions on Wireless Communications*, 5(8):2164–2174, 2006.

[105] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath. Networked mimo with clustered linear precoding. *IEEE Transactions on Wireless Communications*, 8(4):1910–1921, 2009.

[106] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Transactions on Communications*, 43(12):2869–2873, 1995.

[107] E. Hyytia, P. Lassila, and J. Virtamo. Spatial node distribution of the random waypoint mobility model with applications. *IEEE Transactions on Mobile Computing*, 5(6):680–694, 2006.

[108] H. Zhou, S. Mao, and P. Agrawal. Approximation algorithms for cell association and scheduling in femtocell networks. *IEEE Transactions on Emerging Topics in Computing*, 3(3):432–443, 2015.

[109] https://www.gnu.org/software/glpk/.

[110] 3GPP Technical Report 36.872. Small cell enhancements for e-utra and e-utran—physical layer aspects. *V12.1.0 (2013-12)*, 77(2):1213–1237, 2014.

[111] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda. Downlink packet scheduling in lte cellular networks: Key design issues and a survey. *IEEE Communications Surveys Tutorials*, 15(2):678–700, 2013.

[112] https://github.com/shankarisi1988/skg_simulator_lte. 2019.

[113] S. Yang, H. Zhou, and Y. Qin et al. Ship: Cross-layer mobility management scheme based on session initiation protocol and host identity protocol. *Telecommunication Systems*, 42:5–15, 2009.

[114] A. H. Zahran, B. Liang, and A. Saleh. Signal threshold adaptation for vertical handoff in heterogeneous wireless networks. *Mobile Networks and Applications*, 11(4):625–640, 2006.

[115] Z. Hanzaz and H. D. Schotten. Analysis of effective sinr mapping models for mimo ofdm in lte system. In *9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1509–1515, 2013.

[116] J. Francis and N. B. Mehta. Eesm-based link adaptation in ofdm: Modeling and analysis. pages 3703–3708, 2013.

[117] https://github.com/shankarisi1988/5g_hetnet_handoff_skg. 2019.

[118] S. Fu, L. Ma, M. Atiquzzaman, and Y. J. Lee. Architecture and performance of sigma: a seamless mobility architecture for data networks. In *IEEE International Conference on Communications*, volume 5, pages 3249–3253 Vol. 5, 2005.

[119] N. Ahmed and N. Rikli. A qos based ahp algorithm for the vertical handover between heterogeneous wireless networks. In *9th IEEE-GCC Conference and Exhibition (GCCCE)*, pages 1–6, 2017.

[120] The analytic hierarchy process, https://www3.diism.unisi.it/ mocenni/note_ahp.pdf.