Chapter 1
Introduction

For the last few decades (1970-2016), information society has undergone a radical transformation from conventional radio, television and telephony system to smart internet enabled communication devices. The main driving force of such society is Information and Communication Technology (ICT) which promises communication, content and computing [1]. But the evolution of internet had a predominant impact on classical ICT. In this era of ICT, there are massive advancements in the field of Internet and Mobile technologies. Today, Internet of Things (IoT) provides several services to various machines which are connected to offer machine-to-machine (M2M) communication [2]. People always want to be connected with the latest information anytime and anywhere. ICT and IoT are engrossing a lot of attention due to the increase in mobile users. The fast-changing network topologies, networking technologies, user requirements, services and application types, etc., are all driving the need for an all-IP networking system that integrates various types of networks, providing a ubiquitous and seamless networking ambience. As a matter of fact, the migration to IP communications has already started its march in industry much time earlier.

Fourth generation communications system (4G) which is also referred as Next Generation Networks (NGN) is targeting at realizing an all-IP based packet switched system which integrates almost all the existing 2G, 3G and 4G technologies [3]. Also, better signals, good coverage, higher data rate and network capacity, better QoS assurance and seamless handoff across heterogeneous networks, etc., are the main objectives of the 4G wireless communication standard. The former wireless generations, i.e., from 1G to 3G were planned for cellular telephone access and provided rich communication experiences, but ‘Beyond 3G’ or 4G or Next generation networks is focusing on heterogeneity of networks. Each network has a different set of capabilities and satisfies different type of user requirements, therefore combining them all leads to Heterogeneous Networks, which aims at providing better experience, services and connectivity. This NGN
focuses on a concept of Anytime, Anyhow, Anywhere (AAA) providing best services by offering ‘Always Best Connectivity’ (ABC) [4].

Different types of wireless networks will be deployed and integrated to offer better coverage, connectivity and reliability. Such a heterogeneous network will ensure to provide optimal QoS and QoE to home and visiting customers.

So with the vision of providing ‘always best connectivity’ to our customers, there is a need to develop a robust architecture which can provide better coverage and connectivity. The paradigm shift from homogeneous network to heterogeneous network is an ideal solution for providing seamless connectivity, increased capacity, traffic offloading and ubiquitous coverage. The key domains of Next Generation Heterogeneous Networks (NGHN) are Internet of Things (IoT) [5], Internet of Services (IoS), ambient intelligence [6], cloud computing, data analytics, mobile computing, ubiquitous coverage and seamless connectivity. And this all is possible if our mobile device is best connected to the Internet all time.

Hence, our focus is on developing a robust mobility framework which can provide uninterrupted services to the users without the forced termination and call dropping. To facilitate persistent communication, efficient handoff algorithm is required to manage the mobility among different wireless access networks to offer constant connectivity.

1.1 Introduction to Wireless Communication

The term “Wireless” which means ‘no wires’ is referred to as a communication technology, in which signals are carried through radio, infrared and over a distance without using wires to connect different communication devices such as Laptops, Pagers, Cell phones, Portables PCs, Computer networks, Satellite systems and Personal Digital Assistants (PDA). Data can be transferred through a wireless network in three main components: radio signals, header information and network framework. These three components work independently of the each other, so need to be defined separately.
The radio signal functions of the physical layer which communicates with the network interface to receive and transmit signals. Header information works with the higher layers. The network framework includes the wireless network routers, interface and adapters. It manages service management, handoff management and performance management and controls the Access Point/Base Station (AP/BS) which sends and receives the radio signals. Figure 1.1 refers to the various generations of wireless from 1G to 5G.

Figure 1.1: Wireless Generations

Wireless LANs offers employees to use the internet and remain connected to their corporate data, statements and emails even whenever they are away from their office. It is a flexible, easily available and trustworthy data communication system, implemented while connected to the backbone as wired. Wireless transmission and reception takes place over the air, as a medium using Radio Frequency (RF) technology.

1.2 Overview of Future Heterogeneous Wireless Networks

Last few decades have witnessed tremendous growth in Internet and Mobile industry. Due to this, mobile broadband traffic has increased drastically and operators are offering best services to their customers to maintain a better grade of service. Further to successfully execute latest projects on ambient intelligence, social media networking, e-commerce and m-commerce etc. requires seamless connectivity with huge data rates [7]. As we move from generations to generations (1G to 4G), the main aim is to provide better services to mobile users. It is generally believed that networks Beyond 3G i.e. 4G or Next Generation Wireless Networks (NGWN) will furnish uninterrupted services at higher data rates with no or negligible delay to make the vision of “ubiquitous computing” become a reality.
This can be done through Heterogeneous networks, which is an integration of varied networks laid down across complete targeted area and users can travel freely across these multiple Radio Access Technologies (RATs) such as Long Term Evolution (LTE), Wireless LANs (WLANs), Worldwide Interoperability for Microwave Access (WiMax), Universal Mobile Telecommunications System (UMTS), WiBro, etc., for consistent internet connection. Figure 1.2 illustrates Heterogeneous Wireless Network in which a user can roam freely in a city in which different types of cells are deployed. A high power node called macro cell is used to act as an umbrella cell and low power nodes such as Micro, Pico and Femto cells are used to provide coverage to targeted areas to prevent dead spots.

![Heterogeneous Network](image)

**Figure 1.2: Heterogeneous Network**

Low power nodes could easily be added to increase coverage and capacity, decrease attenuation in dead spots (devoid of signals) to increase consumer demand. They are used to fill in areas where signals cannot be penetrated by the macro network – basement, outdoors and indoors [8]. They promise high network performance and service quality by addressing to various issues like load balancing, multi-homing, mobility management and resource allocation [9].
Heterogeneous network planning works well with the co-operation and co-ordination of high and low power nodes of random frequencies. Two different power capacity cells when combined together, they provide solutions for traffic handling, packet offloading and performance upgrading. The heterogeneous wireless networks and their IP based solutions make them more and more appropriate for an ample range of applications in different areas such as public safety, healthcare, smart farming, environment monitoring, etc. Figure 1.3 illustrates the use of different types of cells to provide coverage in a city.

![Wireless Heterogeneous Network Scenario](image)

**Figure 1.3: Wireless Heterogeneous Network Scenario**

The deployment of macro cell is used to provide wider coverage as compared to Pico or Femto cells, which are employed to cover small areas or regions.

### 1.3 Cellular Concepts

The basic purpose of the Personal communication system is to provide maximum coverage and to raise the value of QoS at all. Earlier the area to be deployed to the Internet and mobile facility used a single antenna which was high powered and provided coverage to larger area but fail to achieve frequency reuse. This makes it clear that the limited frequency band can provide services to a limited number of users. To use this limited spectrum by a number of users without causing spectral congestion, some new design methodology is needed and the answer is cellular
concept. The cellular concept introduced different types of cells, such as Pico, Femto, Micro and Macro which differ from each other on the basis of the coverage and capacity. Table 1.1 gives the detail about their coverage and signals transmitted by them.

<table>
<thead>
<tr>
<th>Type of node</th>
<th>Transmitted Power (dBm)</th>
<th>Antenna Gain dB</th>
<th>Range</th>
<th>Cable loss dB/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro</td>
<td>46</td>
<td>~12-15</td>
<td>30 Kms</td>
<td>0.5</td>
</tr>
<tr>
<td>Pico</td>
<td>23-30</td>
<td>0-5</td>
<td>200 m</td>
<td>0.2</td>
</tr>
<tr>
<td>Femto</td>
<td>&lt; 23</td>
<td>2</td>
<td>10 m</td>
<td>0</td>
</tr>
</tbody>
</table>

Putting low and small power antennas in place of a single high power antenna is again a part of a cellular design. Cellular concept has two main laws:-

i. Neighbouring cells are using totally different RF channels causing no interference.

ii. Frequency reuse is possible for the distant cells.

1.3.1 Cell Structure

The base station is situated almost in the centre of the cell providing equal power and range in all directions as shown in Figure 1.4. The radio coverage by one base station is referred as a cell. In a cellular system, the shape of a cell is an important feature which is circulating in an ideal situation as shown in Figure 1.5 (a). There are many constraints imposed by natural hindrances like hills, man-made structures like buildings, towers, etc., which affect the ideal shape of the cell. The actual shape of the cell may not be either a circle but little distorted along the edges as shown in the Figure 1.5(b) [10]. But for analysis and study of Wireless and Mobile Communication system, an appropriate model for cell shape is needed, e.g., hexagonal, rectangular or triangular as shown in Figure 1.5(c).
Why hexagonal and not circular?

There are two main reasons for taking hexagonal shape:

i. To fill the gaps which are created due to a circular shape. These gaps are called dead spots because they are devoid of signals or have negligible radio coverage as shown in Figure 1.6.

ii. No doubt that due to overlapping, there is a complete coverage, but when the Mobile Station (MS) is in the common area, it is difficult to decide that which cell may provide the service and may cause interference.
1.3.2 Frequency Reuse

Traditional radio communication system was facing problems related to limited service area capability and inefficient spectrum utilization causing congestion and delay in communication. But the current wireless and mobile communication system has overshadowed these problems through an efficient usage of available spectrum [11]. This is achieved through the use of smaller cells and reuse of frequency after a safe distance from the current cell. Frequency reuse is the procedure of allocating same frequency to cell adjacent to its neighbour to avoid co-channel interference, i.e. the same frequency can be used by the distant cells. Therefore a group of cells which constitute to have a different frequency is called a cluster. Main attributes of frequency reuse are:

i. Covering a large area through small cells.
ii. Increasing system capacity.
iii. Providing efficient spectrum utilization. A regular hexagonal pattern results in obtaining optimal area coverage and efficient spectrum utilization.

Cells with same colour constitute a cluster as shown in the Figure 1.7. In a cluster, all the cells are having different frequency; but in a different cluster, the frequency can be reused [12].

![Figure 1.7: Frequency reuse](image)

1.3.3 Channel Assignment Strategies

The aim and purpose of frequency reuse strategy is to grant appropriate spectrum utilization without causing any congestion and minimizing co-interference between
adjacent cells. Channel assignment schemes help in achieving these objectives more efficiently. The channel can be assigned to a cell in two ways:

i. Static/Fixed channel assignment
ii. Dynamic/Random channel assignment

i. **Static Channel Assignment:** A fixed set of voice channels is assigned to an access point or base station. When a request for a radio resource is made, then the call is processed; if free channels are available otherwise a call is blocked or it fails. Such calls which are blocked due to non-accessibility of channels are called blocked calls. Sometimes a request for existing calls is turned down because the channel is not free which is called forced termination. Forced termination is less desirable than call blocking.

ii. **Dynamic channel Assignment:** These schemes are more appropriate to allocate channels efficiently. It serves the following objectives:

a. Reducing call blocking probability, by dynamically assigning the channels to a cell, but also putting a check on the number of channels being allocated because there is a limit on the frequency band being offered to each cell, and thereby; there is an upper limit to the number of channels allocated per cell.

b. Another assignment can be on the basis of the amount of traffic handled by each cell.

c. Another alternative arrangement can be dynamic assignment of channels on the basis of current demand.

In dynamic channel allocation, the central server maintains all the available channels. The current demand is assigned channels from this available channel. There is no permanent allocation of channels. When a call is made, then the request is forwarded to the Main Switching Centre (MSC). Then, MSC checks if the channels are available, then it will allocate it to the requested call.
1.4 Path Loss Models

The wireless radio channel is vulnerable to many impediments due to its dynamic nature. It has trouble penetrating through the glass, mountains, hills, huge towers, etc. Hence, these hindrances in the surroundings obstruct the signal and it attenuates.

The distance covered by the signal till it reaches the receiver decides the amount of attenuation. This distortion in signal strength transmitted by the antenna with propagating through the unpredictable channel is termed as path loss [13]. Shadowing is another term which is used in conjunction with path loss, but it is caused owing to hindrances between the transmitting antenna and receiver which dissipate the signal strength due to absorption, reflection and diffraction. Many path loss models have been developed which considers the nature of the complex wireless channel taking into account urban, sub-urban as well as open area.

The promulgation of radio wave of urbanized district is strongly affected by the size of obstacles, chaotic behaviour of signals and density of buildings which results in slow or fast fading. For the sake of propagation study by the researchers, an environment area for most of the time is employed qualitatively by making use of terms corresponding to dense urban, urban, suburban and rural areas. 'Dense city' discipline mostly defined as the area covered by tall buildings, towers, well constructed offices, workplace blocks and different business buildings, whereas suburban areas comprise of 'suburb' or residential construction on outskirts surrounded by lawns or parks. The open or 'rural' area defines open farmhouses with sparse structures, small wooden constructions and small houses with a low density population. To study path loss models, six key factors are used in classifying land utilization[14] [15]:

i. The specific area covered by the buildings
ii. Height of the building and transmitter
iii. Building size (area covered by buildings)
iv. Nature of area (urban, sub-urban or open)
v. Kind of trees or vegetation nearby
vi. Land density
These models are categorized into three categories: deterministic, empirical and stochastic models. But in this thesis, only three empirical path loss models have been considered which include free space, okumara and hata path loss models.

1.4.1 Free Space Path Loss Model

This model is used to detect signal strength of radio signals present in a radio system [16]. The model is used as a basis for learning propagation of radio signals in real life environments.

**Computation of free-space path loss:**

The formula for computing path loss in free space is shown in eq 1.1. The path loss is relative to “the square of the distance between the transmitting and receiving antenna, but the signal level is proportional to square of the frequency” [17].

\[
F_{SPL} = \left(\frac{4\pi d}{\lambda}\right)^2
\]

where:

- \(F_{SPL}\) = path loss in Free space
- \(d\) = distance in metres of the receiver from the transmitter
- \(\lambda\) = signal wavelength (metres)

1.4.2 OKUMURA model

This model can be employed in urban, suburban and open regions and is built originally based on the data gathered from Tokyo, Japan [18]. The model is appropriate for urban architectures but cannot be employed for tall blocking frameworks. “Open areas” refer to zones where there are no tall trees/buildings in the path, but consist of a plot of land cleared. Suburban area covers a village/highway scattered with trees and houses. “Urban area categories are built up city or large town with large buildings and houses with two or more storey or larger villager with close houses and tall, thickly grown trees”.

11
Computation of Okumura path loss:

The path loss is computed using the equation 1.2\[18][19][20] as

\[
\text{PL}_m (dB) = L_{FS}(d) + \text{Att}_\text{mu}(f,d) - G(h_t) - G(h_t) - G_{\text{AREA}}
\]  
(1.2)

- \(\text{PL}_m\) = median of path loss
- \(L_{FS}(d)\) = free space propagation path loss.
- \(\text{Att}_\text{mu}(f,d)\) = median attenuation relative to free space
- \(G(h_t)\) = gain factor for base station antenna height
- \(G(h_t)\) = gain factor for mobile antenna height
- \(G(h_t)\) = gain factor for mobile antenna height
- \(G(h_t)\) = 20log\((h_t / 3)\) for \(10m > h_t > 3m\)
- \(G(h_t)\) = 10log\((h_t / 3)\) for \(htm <= 3m\)
- \(G(h_t)\) = 20log\((h_t / 3)\) for \(10m > h_t > 3m\)

1.4.3 HATA model

Hata formulated a set of experimental relationships which explains the graphical information given by Okumura. The formulas are restricted to a specific set of input parameters and is relevant only on quasi-smooth terrain [21][22].

The expression and its applicability can be done as follows:

Carrier Frequency: \(150 \text{ MHz} \leq f_r \leq 1500 \text{ MHz}\)
Antenna Height Base Station (BS): \(30 \text{ m} \leq h_b \leq 200 \text{ m}\)
Antenna Height Mobile Station (MS): \(1 \text{ m} \leq h_m \leq 10 \text{ m}\)
Transmission Distance: \(1 \text{ km} \leq d \leq 20 \text{ km}\)

\(A + B \log_{10} (dt)\) for urban areas

\(\text{PL}_p (dB) = A + B \log_{10} (dt) - C\) for suburban area

\(A + B \log_{10} (dt) - D\) for open area

where:

- \(A = 69.55 + 26.161 \log_{10} (f_r) - 13.82 \log_{10} (ht_b) - s (ht_m)\)
- \(B = 44.9 - 6.55 \log_{10} (ht_b)\)
- \(C = 5.4 + 2 \left[ \log_{10} \left( \frac{f_r}{28} \right) \right] 2\)
- \(D = 40.94 + 4.78 \left[ \log_{10} (f_r) \right] 2 - 18.33 \log_{10} (f_r)\)
According to the prevailing and the requirement of the work, many more path loss models can be incorporated, but the scope of this thesis limits for these three models and rest are not a part of this thesis.

1.5 Mobility Management

The integration of the Internet and Personal Communication Service (PCS) has attracted many customers, thus increasing the number of mobile subscribers. The use of smart handheld devices makes customers more flexible, scalable and mobile and they prefer to use the internet while away from their desk. Here comes the role of Mobility Frameworks Management, which comes forth as one of the most significant and thought-provoking subject for the industry. Mobility management allows seamless connectivity to users by serving them in two modes. First, through location awareness which enables to locate the position of the user and maintaining databases [23] such as HLR (Home Location Register) and VLR (Visited Location Register) (i.e. Location management), and Second, maintaining an uninterrupted service delivery while moving (handoff management).

1.5.1 Location Management

Location management strategy was proposed in early nineties by [24], which studied traffic load and terminal mobility in an European standard GSM “Global System for Mobile Communications”. It deals with the process of locating the mobile when user moves from one location to another. For this, location management process maintains two-tier databases of each place where a mobile user visits. It is used to trace the positions of the mobile user. Location management is done in two steps [25]:

i. Location registration

ii. Call delivery or paging

where, $s (ht_m) = [1.1 \log_{10} \left(fr_c \right) - 0.7] \cdot ht_m - [1.56 \log_{10} \left( fr_c \right) - 0.8]$ for medium or small cities

8.29[log_{10} \left(1.54 ht_m\right)]^2 - 1.1 for large city and $fr_c \leq 200$ MHz

3.2 [log_{10} \left(11.75ht_m\right)]^2 - 4.97 for large city and $fr_c \geq 400$ MHz

Carrier Frequency: $150 \text{ MHz} \leq fr_c \leq 1500 \text{ MHz}$

Antenna Height Base Station (BS): $30 \text{ m} \leq ht_b \leq 200 \text{ m}$

Antenna Height Mobile Station (MS): $1 \text{ m} \leq ht_m \leq 10 \text{ m}$

Transmission Distance: $1 \text{ km} \leq dt \leq 20 \text{ km}$

$A + B \log_{10} \left( dt \right)$ for urban areas

$P_{L_p} (dB) = A + B \log_{10} \left( dt \right) - C$ for suburban area

$A + B \log_{10} \left( dt \right) - D$ for open area

where:

$A = 69.55 + 26.161 \log_{10}(fr_c) - 13.82 \log_{10}(ht_b) - s(ht_m)$ (1.3)

$B = 44.9 - 6.55 \log_{10}(ht_b)$ (1.4)

$C = 5.4 + 2 \left[ \log_{10} \left( fr_c/28 \right) \right]^2$ (1.5)

$D = 40.94 + 4.78 \left[ \log_{10} \left( fr_c \right) \right]^2 - 18.33 \log_{10}(fr_c)$ (1.6)

$s(ht_m) = \left[ 1.1 \log_{10} \left( fr_c \right) - 0.7 \right] \cdot ht_m - \left[ 1.56 \log_{10} \left( fr_c \right) - 0.8 \right]$ for medium or small cities

8.29[log_{10} \left( 1.54 ht_m\right)]^2 - 1.1 for large city and $fr_c \leq 200$ MHz

3.2 [log_{10} \left( 11.75ht_m\right)]^2 - 4.97 for large city and $fr_c \geq 400$ MHz
Location registration involves the updating of HLR and VLR by the mobile user periodically by exchanging few signals via mobile switching centre about its current location to keep location database updated [26]. VLR stores the location which a mobile user is visiting temporarily and HLR stores the home location of the user. There is an interface between the HLR and VLR to share the mobile user’s detail. Once location registration is complete, the second step is called. Based on the information residing in the VLR and HLR, the call is transferred to the user successfully. For the successful execution of location management process, following issues must be kept in mind:

i. Reducing latency for maintaining consistency
ii. Considering the densification of the area
iii. An increase in the mobile subscribers has increased the traffic load so while tracking Mobile Unit (MU), there can be a significant increase in the network load which needs to be monitored.
iv. Updating HLR and VLR as mobile user moves in a geographical area is a four way communication which should progress through a secure channel.
v. Avoiding call blocking and prioritizing calls during call delivery procedure

1.5.2 Handoff Management
The continuation of the services provided by the wireless environment is one of the most significant characteristic of PCS. To achieve continuity, maintaining of reliable and good links between the mobile user (laptop, mobile phone) and Access point or base station is must which uses handoff mechanism. “Handoff or Handover occurs when a MS switches from currently connected AP/BS on one network to another base station of same or different network, when network conditions or performance of the MS degrades” as shown in Figure 1.8. For example, when a MS does not have sufficient bandwidth to support Voice over IP (VOIP) services, so it will search for a better network or when their signal strength is too low due to call continuity is not possible.
1.6 Handoff

When a handoff process is initiated, it keeps many network resources busy. During the process, there is no service available in few seconds on the MS because of switching between networks. This delay in service continuity is called handoff delay. There are many kinds of handoff delay which occurs during the complete procedure and are discussed in the section. It is not an easier task because it has to consider mobile, user, application and network requirements.

![Handoff Diagram](image)

**Figure 1.8: Handoff Process**

Hence, varied types of methods are being proposed and researches are adopting mobility management as an exciting topic for research and for optimization of performance and results. Even use fuzzy logic and neural networks are being employed to choose the best AP or base station or best network.

1.6.1 Handoff Strategy

With the announcement of heterogeneous networks, the need of handoff in a ubiquitous environment has increased.

Accordingly, Handoff can be classified into the following types:

i. **Intracellular handoff**: During this handoff, link transfer is made between two time slots or channels or on the basis of frequency within the same BS and within the same cell.
ii. **Intra BSC handoff**: When a handoff takes place between two different APs, but both connected to the same Base Station Controller (BSC). Figure 1.9 shows the various strategies of handoff process. When a car moves from AP₁ to AP₂ connected to BSC-1 and then Intracellular handoff takes place.

iii. **Inter BSC handoff**: This kind of link transfer between two base stations linked to the distinctive BSC on same MSC. In Figure 1.9, when a car moves from AP₁ of BSC-1 to AP₂ of BSC-2, then Inter BSC handoff takes place as both BSCs belongs to same MSC.

iv. **Inter MSC handoff**: This kind of link transfer between two base stations linked to the distinctive BSC on different MSC. In Figure 1.9, when a car moves from AP₂ of BSC-2 of MSC-1 to AP₁ of BSC-3 of MSC-2 then Inter MSC handoff takes place as both BSCs belongs to different MSC.

v. **Intersystem handoff**: In this approach bond is transferred from two BS on different BSC to different MSC belonging to different PCS networks.

### 1.6.2 Handoff Type

Handoff can be classified into four broader categories as shown in Figure 1.10. The detail of each category is explained as below:
1. **Accessible technology**: There are two types of handoff depending on access technology:

   a) **Horizontal handoff** happens when a mobile user switches from one cell coverage area to another cell on the same network, e.g., homogeneous systems such as Code Division Multiple Access (CDMA) and Global System for Mobile communications (GSM).

   b) **Vertical handoff** proceeds when a user transfers the control from one AP on one network to another AP of different networks. Both the networks are a part of heterogeneous networks, which comprises of varied types of Radio Access Technologies. They are categorized as:

   i. **Upward vertical handoff** happens when cell user transfers the control of small coverage cell to wider coverage cell of different networks.

   ii. **Downward vertical handoff** is vice-versa of upward handoff in which control is transferred from wider coverage area to lower, i.e., in the reverse direction.

   Apart from these types of handoff, which occurs due to radio link quality, handoff can also be triggered on performance or user basis to support efficiency in wireless data services. In this perspective, vertical handoff deals with the heterogeneities of the interconnected wireless networks [27] and classified as:

   i. **Imperative handoff** occurs when a mobile user performs handoff due to loss of signal strength or signal-to-noise interference. In this, handoff has to take place immediately; in any other case it will result in a lack of connection and discontinuity of services. Further, they are classified into two sub-categories:

      a. **Reactive handoff** occurs according to the information provided by the network to MU regarding availability of connection or loss of signals. It can be further divided into **anticipated handoff or unanticipated handoff**. An **anticipated reactive handoff** takes place when there is an alternative BS / AP is available to which the MS can handoff. Where as in case of **an unanticipated reactive handoff** if the current network is devoid of signal strength and the MS has no other point of attachment (PoA) to which it can perform handoff, then the MS is
enforced to temporarily shut down its ongoing connections and resume them when a new network is available [28].

b. **Proactive handoff** performs handoff while maintaining communications. After the network decides about handoff, then only communication channels are broken off to change or shift to a new spectrum band/BS. User mobility and cell overload are the examples of proactive handoff events.

ii. **An alternative vertical handoff occurs when the mobile user switches due to** better performance than any loss of signal [29]. In this, QoS parameters play a key role such as available bandwidth, delay, throughput, etc. The handoff procedure can be classified as hard handoff or soft handoff.

2. **On the basis of layered approach:** According to the protocol layer suite [25], handoff can be classified as:

i. **Layer 2 handoff** occurs at Datalink Layer when an MS experiences loss of signals and in order to enjoy incessant connection; it changes its PoA from one AP to another on the same subnet. Based on the information received from the lower layers (physical Layer) about Received Signal Strength Indicator (RSSI), data rate and Signal to Interference ratio (SINR), networks can take a decision of switching to scanned AP/BS. It releases the association with the preceding AP and builds a fresh connection with the scanned AP. It is restricted to the same network and does not seek help from above higher layers for carrying out the handoff procedure.

ii. **Layer 3 handoff** takes place on Network Layer and occurs on different subnets. Network layer contributes in managing handoffs between two BS/APs of dissimilar subnets. In this case, MS can switch between networks belonging to different Base Switching Centre (BSC) or Mobile Switching Centre (MSC). Hence, to carry out Layer 3 handoff, cross layer approach contributes efficiently to complete the handoff procedure successfully.

Handoff process involves the cross-layer mechanism. In order to manage and execute handoff process; more than one layer involvement is required e.g. in order to achieve network layer handoff; lower layers and upper layers contribute and participate in providing all necessary information to the network layer. On the
receipt of this information like about signal strength, SINR or link status, in advance, the system gets ready for executing handoff without causing call dropping or call blocking [25].

3. **Depending on the technology, active connections and signalling process:** Based on this, handoff can be divided into three categories:

i. **Hard handoff** takes place when the association with the current AP/BS is broken before making any new connection. After breaking with the old, the new link or connection is made with the selected base station (break before make). This results in a slight delay in providing services to the customers affecting the QoS. Hard handoff is used by GSM (Time Division Multiple Access) and General Packet Radio Service (Frequency Division Multiple Access). Further, it can be divided into two categories:
   
a. **Inter Carrier frequency handoff** occurs when the carrier frequency of the latest scanned base station is different from the preceding one.
   
b. **Intra Carrier frequency handoff** occurs when the carrier frequency of the latest BS is similar as the earlier one.

ii. **Soft handoff** occurs when the connection with the scanned candidate network or BS/AP is recognized first and then the connection with the preceding BS is relinquished. This is done to provide seamless connectivity to users (make before break). It is used by Code Division Multiple Access (CDMA) e.g. HSUPA and Wideband CDMA (WCDMA). Soft handoff is better than a hard handoff and employs Fast Base Station Selection mechanism (FBSS).

During FBSS, each MS maintains an active set of BS/APs in which the list grows, when the BSs having the signal strength above the certain threshold are added and the list shrinks when the BSs are removed as its signal strength falls below the minimum value. But MS seamlessly remains connected with one base station to maintain session continuity. This base station is called the Anchor Base station. In softer handoff, which is a special case of soft handoff, MS is connected to two dissimilar sectors of the similar (connected) cell. It usually takes place when a cell sectoring is done for frequency reuse and increasing capacity.
Figure 1.10: Soft handoff and hard handoff

Mobile user is connected with both APs in soft handoff before break.

Mobile user breaks with previous AP then makes with new AP.
Figure 1.11: Types of Handoff
4. **Type of controlling and assisting entity:** On this basis, handoff is divided into three categories explained as follows:

i. **Network Controlled Handoff (NCHO)** occurs when the principal control of the handoff procedure is kept on the network. The network performs all the necessary measurements and performs handoff at MS.

ii. **Mobile Controlled Handoff (MCHO)** occurs when the MU has the primary control of the handoff process. In this MU controls the handoff process and informs it to the BS/AP if handoff required.

iii. **Mobile-Assisted-Handoff (MAHO)** occurs when the mobile station based information and preferences are forwarded to the base station and base switching centre. In this type of handoff is executed by the network, but with the assistance of the mobile station. While, in a Network-Assisted-Handoff (NAHO), the BSC with the help of BS itself collects the required information and passes it to the MU for handoff decision.

1.6.3 **Handoff Metrics**

Handoff metrics are used to study the current state of MU and Network. They can be classified into five categories:

i. **Technical or Network related parameters:** RSSI, SINR, BER, Network load

ii. **QoS parameters:** Bandwidth, Throughput, delay, Jitter, Packet Loss

iii. **Mobile node parameters:** velocity, direction

iv. **User preferences:** cost, security, data rate.

v. **Miscellaneous:** history, response time, estimated time, Round Trip Time (RTT)

It is necessary to study these parameters to decide what is to be measured and what is required for execution of handoff. Table 1.2 summarizes those parameters which need to be measured during handoff initialization and decision making phase. It also provides those parameters which must be output after the handoff process takes place to measure its efficiency.
Table 1.2: Parameters to be considered for analysis and measuring performance

<table>
<thead>
<tr>
<th>Which metrics can be measured or input</th>
<th>Handoff metrics measured during handoff to check the handoff algorithm performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSSI</td>
<td>Load</td>
</tr>
<tr>
<td>SINR</td>
<td>Jitter</td>
</tr>
<tr>
<td>BER</td>
<td>Bandwidth</td>
</tr>
<tr>
<td>Distance</td>
<td>Cost</td>
</tr>
<tr>
<td>Velocity</td>
<td>Security</td>
</tr>
<tr>
<td>Network Coverage</td>
<td>Jitter</td>
</tr>
<tr>
<td>Delay</td>
<td>History</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>Direction</td>
</tr>
<tr>
<td>Handoff delay</td>
<td>Handoff Call blocking</td>
</tr>
<tr>
<td>New Call Blocking</td>
<td></td>
</tr>
<tr>
<td>Outage Probability</td>
<td></td>
</tr>
<tr>
<td>Number of Handoffs</td>
<td></td>
</tr>
<tr>
<td>Packet Loss</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.12 explains all the types of parameters for the entire handoff process. It is difficult to realize all the metrics in one algorithm so category-wise demarcation has been done. Depending upon what is required in a process, parameters can be chosen.

![Handoff metrics diagram](image)

Figure 1.12: Handoff metrics

Table 1.3 discusses each of the metric and gives a proper definition of each along with the classification.

Table 1.3: Handoff Metrics [30]

<table>
<thead>
<tr>
<th>Source</th>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>RSSI</td>
<td>This criterion is most widely used. Especially in case of Line-of-sight (LOS) and limited interference systems, there the RSSI indicates the signal quality, and is accurate. As the distance between the increases, the RSSI value starts decreasing. As it falls below the minimum threshold, handoff is required.</td>
</tr>
<tr>
<td><strong>SINR</strong></td>
<td>SINR is a metric associated with noise, interference, and signal strength. It is defined as the ratio of signal to interference ratio. Its unit is also dBm but is different from RSSI. Lower value of SINR means low signal strength due to high interference.</td>
<td></td>
</tr>
<tr>
<td><strong>Data Rate</strong></td>
<td>The rate at which the data is being transferred between a peripheral device and the computer is called data rate and it is measured in bytes per second (bps or Mbps).</td>
<td></td>
</tr>
<tr>
<td><strong>BER</strong></td>
<td>The bit error rate is defined as the ratio between the number of bits in which error incurred during the transfer and the total number of transferring bits during a calculated time interval.</td>
<td></td>
</tr>
<tr>
<td><strong>Retransmission</strong></td>
<td>It is the process in which packets are retransmitted in case of lost frames, damaged frames or lost acknowledgement.</td>
<td></td>
</tr>
<tr>
<td><strong>Bandwidth</strong></td>
<td>Total capacity of the channel or frequency band available to or used by the communications channel [31].</td>
<td></td>
</tr>
<tr>
<td><strong>Packet Loss</strong></td>
<td>Measures the average packet loss rate within the network.</td>
<td></td>
</tr>
<tr>
<td><strong>Delay</strong></td>
<td>Interval between the packet arrival at the sender's end and the packet reception time at the recipient.</td>
<td></td>
</tr>
<tr>
<td><strong>Latency</strong></td>
<td>Time elapsed between the packets send and received during handoff, there are many messages which are exchanged among BS, MS and BSC which adds up to cause latency. It is the mean delay which is caused during handoff process.</td>
<td></td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>The quantity of information or packets successfully transferred from a resource to a target is called throughput. It is measured in bps (bits per second).</td>
<td></td>
</tr>
<tr>
<td><strong>Velocity or speed of mobile station</strong></td>
<td>Velocity is an important location based handoff metric. It is defined as the rate at which the mobile user is travelling with time. Numerous schemes use velocity to take decision about the handoff especially in case of overlay networks.</td>
<td></td>
</tr>
<tr>
<td><strong>Direction</strong></td>
<td>It is the direction in which the mobile node is progressing.</td>
<td></td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>Mobile users have become very smart and want full value for what they pay. They have their set of preferred qualities in a network related to quality, availability, reliability and maintainability. These sets of their preferences are called User preferences. For e.g. users preferred more secure network available at lower cost.</td>
<td></td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>It is the process of protecting the network resources from unauthorized access to maintain confidentiality or integrity of the transmitted data. So during handoff network selection, priority may be given to a more secure network than the previous.</td>
<td></td>
</tr>
<tr>
<td>Battery Power</td>
<td>It is a sort of an electric power required for handoff process. Today in smart phones, if a mobile is running short of battery, then mobile can switch over to battery saver which consumes less battery. In case of handoff, a mobile user when running low for battery power, then it may switch to another network that would consume less battery [32].</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>It refers to the type of application run by user e.g. Skype, Online games, Facebook, etc. Each application requires different bandwidth.</td>
<td></td>
</tr>
<tr>
<td>Other Network related parameters</td>
<td>Network Load</td>
<td>Network load is defined as the traffic or number of users connected to the network. If the load on the network is more; then it leads to overloading problems and hinders the performance of the network. It represents the ratio of allocated bandwidth of the total bandwidth.</td>
</tr>
<tr>
<td>Coverage</td>
<td>The geographic area where the station can communicate</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>This attribute allows memorizing the overall score given to the available alternative.</td>
<td></td>
</tr>
<tr>
<td>Response Time</td>
<td>It is the difference between the requested call and call accepted</td>
<td></td>
</tr>
<tr>
<td>Handoff performance evaluation metrics</td>
<td>Handoff Call blocking</td>
<td>It is defined as the probability that a handoff cannot be completed successfully due to unavailability of sufficient resources.</td>
</tr>
<tr>
<td>New Call Blocking</td>
<td>It is defined as the probability of a new call being blocked due to insufficient signals.</td>
<td></td>
</tr>
<tr>
<td>Outage Probability</td>
<td>Due to dynamic and rapid transition in real channel characteristics, there is a probability of failing to achieve the required level of output SINR threshold value.</td>
<td></td>
</tr>
<tr>
<td>Number of handoffs</td>
<td>Number of times mobile station switches from one AP/BS to another due to weak signal strength or when performance degrades.</td>
<td></td>
</tr>
<tr>
<td>Ping pong rate</td>
<td>Ping pong occurs when a mobile station switches between two APs due to fluctuation in signal strength or when no clear indication is there about user movements near the cell edges.</td>
<td></td>
</tr>
</tbody>
</table>

1.6.4 Phases of Handoff procedure
Handoff process is designed to work in three phases:

i. Handoff Initiation phase

ii. Handoff Decision phase

iii. Handoff authentication and re-association phase
1.6.4.1 Handoff Initiation

Handoff initiation is used to trigger the handoff to decide that whether a handoff must take place or not. RSSI is the key metric.

RSSI indicates the signal power being received by the receiver radio after the antenna and possible cable loss. The higher is its value, the stronger is the signal. The value of RSSI closer to the base station is 0; it decreases as the MU moves away from the BS. It is measured in dBm. Handoff must be initiated before the signal fades away so these are the following methods to initiate a handoff:

i. Choosing between two neighbouring RSSI values  
ii. Initiating using RSSI with Threshold  
iii. Initiating using Hysteresis  
iv. Initiating using both hysteresis and threshold  
v. Using dwell timer

i. Choosing strong RSSI

As the mobile user starts moving around the cell and slowly progresses towards the cell edge, RSSI value starts diminishing. The value of RSSI is measured over time and as it comes in the vicinity of other BS, then signals from both the BS are compared and one with strongest signal is chosen to handoff.

Disadvantage

This increases risk of unnecessary handoffs. When an MU is moving away from one BS and near to another, there comes the point when both BSs have similar RSSI...
value. As the user moves further away such that the RSSI (AP\textsubscript{1}) is less than the RSSI (AP\textsubscript{2}), the handoff is initiated. Moreover, the RSSI value is further vulnerable to shadowing effect and due to random or unpredictable moving pattern of the user; it is expected to cause handoff among the two base stations backward and forward constantly. This is called “Ping-pong effect”. It increases the network load and results in wastage of resources as handoff is activated again and again due to unpredictable user movement.

ii. Initiating using RSSI with Threshold

RSSI with threshold introduces a minimum fixed threshold which is defined as the minimum signal strength required for continuation of services of MS. The range of value varies from network to network, but usually it is negative. This is due to the fact that at base station, its value is 0. But as the node starts moving away from the base station, signal strength decreases to below 0. Usually it lies between -3 dBm to -100 dBm. When the MS is approaching the cell boundary, RSSI goes on decreasing. In this case, when RSSI (AP\textsubscript{1}) becomes less than the threshold, handoff is initiated.

iii. Initiating using Hysteresis

Hysteresis value is defined as the difference or the margin value between two RSSI’s being compared. When the RSSI of AP\textsubscript{1} becomes less than the RSSI of AP\textsubscript{2} with margin difference of hysteresis, handoff is initiated. This is done to eradicate the dilemma of ping-pong during handoff commencement. This can be explained as follows:

$$\text{If } \mu^\text{AP}_r - \mu^\text{AP}_r = \text{hys then}$$

handoff is initiated

iv. Initiating using both hysteresis and threshold

Considering both threshold and hysteresis values while initiating handoff helps in solving ping pong problem. Handoff is initiated when RSSI (AP\textsubscript{1}) is below the threshold and RSSI (AP\textsubscript{2}) is stronger than RSSI (AP\textsubscript{1}) by the hysteresis value ‘hys’.
RSSI based threshold is the lowest tolerable RSS value to support services and call continuation. As the RSSI falls below the threshold, hysteresis margin is checked. When the margin between the two neighbouring signal values is greater than or equal to the hysteresis marginal, initiation takes place. This introduces a short duration gap between handoff to take place which puts calls in queue or dropping list.

1.6.4.2 Handoff Decision phase

In this phase, decisions about the optimal access point on same or different network is taken. Network selection is an integral part of this phase, which depends on certain parameters classified as QoS parameters, Location based, network and user preferences. When a user moves freely into a region, one base station cannot support communication beyond its coverage.

Hence, mobile terminal needs to change its PoA (Point of Attachment) from the old to the new base station. Earlier methods were making this decision based on a single metric, i.e. RSSI (Received Signal Strength Indicator) but today, many techniques like Markov chains, queuing, etc., are being used for providing better network for uninterrupted services. But still to deal with left over issues to make smart and intelligent handoff decisions, new adaptive algorithms are needed.

1.6.4.3 Handoff authentication and re-association phase

Once the optimal network is selected during handoff decision making phase, then link transfer takes place. The link to the previous AP is broken and with new AP of a new network is made for validation. It is called re-authentication process because every time user wishes to connect to new AP, it needs to authenticate its identity for security reasons. It is explained in [33] that IEEE 802.11 supports four types of authentication methods.

Handoff process starts when the MU sends a frame to the selected AP for authentication. If positive acknowledgement is received, then re-association process starts: but if no acknowledgement or negative acknowledgement is received, then re-authentication frame is retransmitted. Once a positive acknowledgement is received by the MS then only re-association process starts. It is actually this phase during
which physical link transfer takes place from old AP to new selected AP after authentication. During re-association, a frame request and response frames are exchanged. In [34] a protocol called Inter Access Point Protocol (IAPP) has been discussed which allows an AP to communicate contextual information with other APs on secure grounds. When a mobile node is near the cell edge or away from its current AP, scanning process starts searching for a new AP with better connection facilities.

There are two types of scanning process: active scanning and passive scanning. Once the new AP is located, the mobile node will send a re-association frame to the new AP which contains its MAC address and Base Station Identifier. The new AP verifies the validity of the old AP by sending its detail to the connected server. The server validates the old AP and sends a reply to the new AP with a secure interface to communicate with old AP. Then both the APs exchange some encrypted messages in Move-response and Move-notify and finally handoff procedure is complete.

After discussing these three phases, it is very important to discuss delays in handoff because every phase introduces some sort of delay which ultimately leads to high latency in complete handoff procedure.

### 1.6.5 Delay in handoff

Handoff is a process which involves the contribution from many layers of the OSI model. The physical layer is used to collect the values of RSSI, SINR and other environmental parameters. Datalink layer executes horizontal handoff. The network layer is responsible for vertical handoff. Many protocols like Session Initiation Protocol (SIP), Mobile Internet Protocol (MIP) which supports handoff sit at session layer and Stream Control Transfer Protocol (SCTP), which helps with multi-streaming during soft handoff, is present in the transport layer.

The complete handoff process is carried out by co-ordination and co-operating entities of different layers of the OSI model. But this does not leave process without
delay. When data has to be passed between different layers many kinds of delay are introduced which may hinder in the QoS of handoff.

i. **Link layer delay**: This layer is responsible for horizontal handoffs. Before the mobile node switches to neighbour base station, scanning is done. During scanning, mobile node scans the nearby base stations if available. It takes some time and sometimes mobile disconnects and then connects to the new base station. This kind of delay which is introduced during scanning and call setup with the new base station is a link layer delay.

ii. **Network layer delay**: This layer is responsible for vertical handoff, which involves obtaining the IP address of new network and then switching the call from one network to another which may take more delay. Further congestion in the network introduces more delay. Thus, this kind of delay is called network delay.

iii. **Transport Layer delay**: Many protocols sit at this layer to monitor and manage the transport layer handoff. Most important protocols in this case are SCTP, which is responsible for multi-homing and multi-streaming during handoff. SCTP makes a new connection with the neighbour network/base station while maintaining connection with the old. As new connection gets stable, it breaks the old connection. This disconnection can be delayed due to re-authentication or re-association phase.

iv. **Application layer delay**: This delay is introduced due to handoff third phase “Handoff authentication and re-association phase”. During this, there is a four way exchange of messages between mobile station, BS and BSC which introduces some delay. Protocols such as IAPP and EAP allow a secure interface to exchange those messages. And the delay incurred during this is called application delay. The MS and the Authentication, Authorization and Accounting (AAA) server causing delay in the handoff.
1.6.6 Ping pong and corner effect

*Ping pong* is a very familiar dilemma in wireless and mobile communication. It occurs when a mobile station switches between two APs to and fro due to rapidly changing signal strengths or user movements. It degrades network performance by increasing network load and wastage of channel utilization. Many algorithms were proposed in the literature to deal with the problem which includes RSSI with threshold and dwell timer. Nowadays many other methods like use of multiple metrics for handoff triggering, estimated time of stay in cell and location updates with previous history are being used to reduce this effect [35][36].

![Figure 1.14: RSSI switch back and forth during Ping-pong](image)

The *corner effect* occurs when a LOS changes to NLOS around the street corner and there is a drastic drop in signal strength by 20 ~ 25 dB over short distances. This is due to increase in the obstruction which breaks the clear line connection between BS and MS.

Grimlund and Gudmundson [37] developed an experimental path loss model for corner effect in which mobile roams across line-of-sight and then experiences a corner effect as it crosses the street corner. This model is given as below:

\[
\mu_{\text{RSSI}} = \begin{cases} 
10 \log_{10} \left( \frac{A}{d \cdot (1+g)^{3}} \right) & d \leq d_c \\
10 \log_{10} \left( \frac{A}{d \cdot (1+d_c/g)^{3}} \cdot \frac{1}{(d-d_c)(1+(d-d_c/g)^{3})} \right) & d > d_c 
\end{cases}
\]

where

- \( g \) = breakpoint in meters
- \( d \) = distance of base station from mobile node
- \( d_c \) = distance of the BS from the corner
1.7 Thesis Organization

The manuscript is organized into seven chapters which focus on the related data. The chapters are organized as follows:

**Chapter 1** is the introductory chapter, which includes preliminaries related to wireless networks and next generation heterogeneous networks. It focuses on the Handoff preliminaries, handoff types, and phases, delays in handoff, handoff metrics and problems during handoff.

**Chapter 2** is the Literature Survey which includes the types of handoff schemes and work done in those fields in literature. It also includes the detail comparison of these handoff schemes.

**Chapter 3** focuses on Aims and Objectives behind this research and includes issues in handoff in heterogeneous networks, motivation and the use of fuzzy logic.
Chapter 4 is proposed Handoff Optimization Algorithms (HOA) and its phases. This chapter focuses on the first phase of HOA – Design and Methodology of Handoff Initialization Phase.

Chapter 5 focuses on Handoff decision making phase of the HOA. This chapter gives the brief introduction of MADM approaches: AHP, FAHP and ELECTRE and explains how they are employed in HOA for network selection.

Chapter 6 is the Results and Performance Analysis. This part presents few scenarios to validate the proposed algorithm, all the experimental models and comparison of HOA with the previous work.

Chapter 7 gives conclusions and future directions. The conclusions of this thesis are presented in this chapter and some future directions which can be derived from this work are mentioned.

Chapter 8 Bibliography