PERFORMANCE EVALUATION OF VERTICAL HARD HANDBOVERS IN CELLULAR MOBILE SYSTEMS

Sahar Bashir Abass, Samah Fageer Ahmed, Shimaa Hyder Makki and Niemah Izzeldin Osman

Department of Computer Systems and Networks, College of Computer Science and Information Technology, Sudan University of Science and Technology, Khartoum, Sudan

ABSTRACT

With the rapid increase of new and diverse cellular mobile services, the overlapping of cells has become typical in the majority of the coverage area of the network. Vertical handovers occur between two layers of cells when a user is switched from one layer to the other. In this paper we investigate the influence of network parameters on vertical hard handover performance in a cell environment. The work considers two layers of cells: a layer of macrocells and a layer of microcells. Handover requests enter the macrocell from neighbor macrocells and from microcells that belong to a different layer. Using Markov chain analysis and simulation we calculate network performance parameters such as mean queue delay, handover dropping probability and channel utilization. We also compare the handover performance for the macrocell and macrocell traffic separately. Our results show the influence of total channels, maximum queue size and handover request arrival rate on handover performance. They also show that when the traffic from each layer is treated with equal priority in the system, the performance of each layer is comparable.

KEYWORDS

Handover Queue Delay, Markov Chain, Performance Evaluation, Vertical Handover

1. INTRODUCTION

The goal of Fourth Generation Systems (4G) is to provide high quality services to the increasing number of users at a reduced cost [1]. Cell overlapping was introduced in mobile cellular systems since 2G systems and is used for many reasons. Using a base station that covers a large area of land is considered an optimum solution for high mobility communication, as is reduces the number of required handovers to maintain communication. In addition, segregating high mobility and low mobility traffic reduces network overhead in terms of channel allocation and handover management. Moreover, the intersection of different mobile generations provides coverage for more recent system (4G and beyond) users where these services are not available yet, ensuring continuous communication.

In a system where macrocells and micro cells overlap, each layer serves users according to some parameter (mobile speed, requested service or channel availability). Handovers occur between the two layers in what is known as vertical handovers. The problem with vertical handovers is not having sufficient resource for users. In some cases, the traffic entering a cell from a different layer might consume all the resources, with the expense of Quality of Service (QoS) for users entering the cell from the same layer.

A number of studies have considered vertical handovers in cellular systems. A few works on handovers in femtocell overlaid macrocell have been described in [2]. The considered handover objectives include throughput, energy efficiency, and reduction in unnecessary handovers,
network load balancing, interference management and network security. The work compares existing handover techniques against the stated objectives. A speed based handover algorithm for macro-femto scenario is proposed in [3]. This study focused on different types of handovers and compared the performance of different handover decision algorithms using simulation. The study showed the mathematical complexity of the most famous approaches for handover techniques and compared them to simulation result.

The survey in [4] discussed various existing handover decision algorithms for the integrated macrocell-femtocell network. A number of handover management algorithms are studied and advantages, disadvantages of proposed algorithms are discussed. The study in [5] simulates femtocell scenarios with features and models compliant with 3GPP specifications using LTE-sim framework. Results show that the increase in the number of femtocells in the system decreases overall performance.

A study, explained in [6] analyzed the performance of handover strategies in femtocell networks under hybrid access mode with the aim of minimizing unnecessary handovers. The evaluation is based on the specific stay time interval ‘T’ and user equipment velocity. The simulation results showed that the proposed algorithm minimized unnecessary handovers decreasing handover probability compared to the traditional strategy.

All of the previous studies consider femtocells, which are indoor miniature cells that serve stationary mobiles. Therefore the mobile speed or the requested service does not play an influential parameter in the handover decision. The main reason for a handover to occur is the mobile exiting the coverage area of the femtocell.

In this paper, we evaluate the performance of hard handover requests that experience a vertical handover between two cellular layers. We consider a two-layer cell environment having a layer of macrocells overlapping a layer of microcells. The handover from a macrocell to a microcell happens due to the change in the speed of the mobile, the limitation of requested resource or the absence of service in the cell. In our work in [7] we show preliminary results for the simulation. Here we consider a more generic model, where overlapping of different cell sizes can be evaluated and therefore exemplifying different mobile generations (2G, 2.5G, 4G, etc). We consider hard handover where each user is connected to at most one base station at all times. Therefore the work does not consider 3G CDMA-based systems that employ soft handover, which requires further investigation. We develop a Markov model and derive analytical equations for different performance parameters. The paper explores the influence of multi-layer traffic on network performance. We compare and validate results from both methods. We consider a more generic model, where hard handover between two overlapped cells is investigated.

The remainder of this paper is organized as follows: Section 2 describes handovers in mobile systems. Section 3 explains the design of the simulation, the input and output parameters and considered system scenarios. Section 4 describes the Markov chain analytical model. In Section 5 we demonstrate by results the vertical handover performance evaluation. Section 6 is a conclusion.

2. **Horizontal and Vertical Handovers**

In cellular mobile telecommunications, the term handover (or handoff) refers to the process of transferring an ongoing call or data session from one channel connected to the core network to another channel [8]. Handovers occur in general when the signal strength between the mobile and the serving Base Station (BS) falls below a certain level due to the mobile distance from the base station or the level of interference. Handovers are classified with respect to systems, frequencies, connections and other criteria. There are two types of handover in general: horizontal handover and vertical handovers.
2.1 **HORIZONTAL HANOVER**

These include all handovers happening between a pair of adjacent cells which may or may not belong to the same system or have the same properties. There are two types of handovers with respect to the number of connections associated with the mobile during the handover process [9]:

2.1.1 **HARD HANOVER**

In hard handover the channel in the source cell is released and only then the channel in the target cell is engaged. Thus the connection to the source is broken before the connection to the target is made. For this reason such handovers are also known as “break-before-make”.

2.1.2 **SOFT HANOVER**

In soft handover the channel in the source cell is retained and used for a while in parallel with the channel in the target cell. In this case the connection to the target is established before the connection to the source is broken, hence this handover is called “make-before-break”. Soft handovers may involve using connections to more than two cells.

2.1.3 **SOFTER HANOVER**

When the soft handover is performed both in the downlink (forward link) and the uplink (reverse link), the handover is called softer. Softer handovers are possible when the cells involved in the handover have a single cell site. In 4G systems the type of handover is hard handover [8], and therefore it is the handover type considered in this work.

2.2 **VERTICAL HANOVER (VHO)**

Vertical handover is a network node that automatically changes its connection type to access a supporting infrastructure. When a computing device could connect to the Internet via two different network technologies, it is automatically connected to the available network. This switching from one network to the other is called vertical handover. Vertical handover enable the phone to select the higher bandwidth at lower costs for networks like wide local area networks. It also provides extended coverage for cellular networks [10]. However, a vertical handover suffers from drawbacks. First, it handles all connections in the same manner. In other words, when all connections are transferred from one interface to another, only one wireless interface (the best one) is used at that moment. In addition, an incorrect handover decision may degrade the QoS or break off the current call.

There are a number of possible handovers in a mobile network, both horizontal and vertical, shown in Figure 1:

1. Handovers moving out of the current macrocell into a neighbor macrocell.
2. Handovers moving from a neighbor macrocell into the current macrocell.
3. Handovers from a microcell into the macrocell.
4. Handovers from the macrocell into a microcell.
5. Handovers moving out of the current microcell into a neighbor microcell.
6. Handovers moving from a neighbor microcell into the current microcell.

In this paper, we consider the handovers labelled 2 and 3 in Figure 1, since they are incoming traffic handled by the resources of the cell of interest (labelled macrocell in Figure 1). As for other handover types, they are handled by resources from other cells.
3. SIMULATION ENVIRONMENT

Following we explain in details the simulation modeling strategy used in the evaluation including the simulation environment, inputs and outputs.

3.1 SIMULATION MODELING

In this paper, we use a discrete event driven simulation to evaluate the vertical handover system [11]. The model of the system evolves over time by an ‘arrive’ or ‘depart’ event and the state variables change instantaneously at separated points in time.

The basic entities in the queuing network model are channels, which represent servers, and handover requests, which represent customers. We use a $M/M/c/c+N$ queue where $M$ describes the Poisson arrival and service process, $c$ are the total cannels and $N$ is the size of the handover queue. The queue is employed to hold handover requests when no channels are available to handle the requests immediately (from the $c$ total channels). ‘Arrive’ and ‘depart’ events are randomly generated and the arrival and service rates follow an exponential distribution. The $c$ channels serve calls from the front of the queue. If there are less than $c$ active calls in the cell, some of the cannels will be idle. If there are more than $c$ calls, additional requests are placed in the queue. The queue is of size $N$, so all requests that exceed the system capacity ($c+N$) are dropped.

3.2 SIMULATION INPUT PARAMETERS

The inputs of the model are:

3.2.1 REQUEST ARRIVAL RATE

A call request in the system is a handover call moving from a neighbor cell into the current macrocell. This call can be entering from a macrocell in the same layer with an arrival rate $\lambda a$, or from a microcell in a different layer with an arrival rate $\lambda i$.

3.2.2 TOTAL CHANNELS

A server here is a channel, which is the resource provided by the cell for the mobile user to send and receive signals for a service (a call for example). There is a total of $c$ available channels in the cell which are fairly granted to macrocell and microcell users.
3.2.3 **SERVICE RATE**

The service rate of a customer at the server \(\mu\) is the number of served customers per unit time. The average service time \(1/\mu\) is the average call duration, which is the average amount of time a caller holds a channel.

3.2.4 **QUEUE LENGTH**

The queue length \(N\) is the maximum number of callers the queue can accommodate.

3.3 **SIMULATION OUTPUTS**

The outputs of the model are:

3.3.1 **CHANNEL UTILIZATION**

The utilization of a channel is the proportion of time the channel is busy, or, equivalently, the average number of customers in service there.

3.3.2 **THROUGHPUT**

The throughput is the average number of served customers per unit time.

3.3.3 **MEAN QUEUE DELAY**

This is the average time a call request waits in the queue.

3.3.4 **DROPPING PROBABILITY**

This is the probability that the call is dropped due to the system reaching its full capacity. (The system capacity is calculated from the queue size \(N\) plus the total channels \(c\)).

3.4 **SYSTEM SCENARIOS**

The system considers two layers of overlapped cells:

1. A layer of macrocells that provides service to mobile users who are moving at a high speed.
2. A layer of microcells that accommodates slow moving and stationary mobile users.

The system considers two types of customers:

1. Handover requests from neighboring macrocells.
2. Handover requests from microcells in a different layer.

There are three possible scenarios in the system:

1. a user request is immediately served
   If there is an available channel, the caller is immediately provided with the channel for continuous communication.

2. a user request is held in the queue
   If all channels are busy, the user request is delayed and placed in the queue until there is an available channel.

3. a user request is dropped
   If all channels are busy and the handover queue is full the user request is dropped.
3.5 SIMULATION FLOWCHART

The flow of events in the simulation is shown in the flowchart in Figure. 2. The simulation starts by initializing the parameters and enter them into the simulation. The next step is accepting a handover requests. The queue is checked whether it is full or not. If the queue is full, then the request will be dropped. In case the queue is empty or not full, the simulation checks if there is an available channel in the cell. If no channel is available, the request will be placed in the queue until a channel is available. Otherwise, the request will immediately be granted a channel. In all cases, statistics are calculated to evaluate the performance of the current request as well as the system in whole, and system state variables are updated. After that, the simulation checks if the end of simulation is reached. If not, it accepts a new request, otherwise, the simulation terminates.

4. MARKOV CHAIN ANALYTICAL MODEL

A Markov chain is a state representation of a system. Each state indicates the number of users in the system. A transition is the rate to move from one state to the next. Figure. 3 below shows the Markov chain of the vertical handover cellular system. An ‘arrive’ process increases the number of handover requests in the system, and it is equally likely to be caused by a request from a macro cell or a microcell. The number of occupied channels is the cell is the number of active callers until all channels \( c \) are occupied. Additional requests result in increasing the number of users in the system, while the total service rate does not exceed the maximum \( c\mu \). The final state indicates that the maximum capacity of the system is \( c+N \).

![Figure.2 Simulation flowchart](image)

![Figure.3 Markov chain](image)
The Markov chain in Figure.3 is solved and a mathematical model is derived for different performance evaluation parameters. Here we show the equations for $P(0)$, $P(n)$ and the $D_q$ which are the probability of no requests in the system, the probability of having $n$ customers in the system and the queue delay, respectively. For the remaining parameters we refer the reader to.

$$
P_0 = \frac{\lambda \mu}{\lambda \mu - \lambda + \mu} \left( \frac{1 - \rho}{\rho} \right) \frac{1}{\rho} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \left( \frac{1}{\lambda} - \frac{\mu}{\rho} \right)
$$

Where $\rho = \lambda \mu$, $\tau = \lambda \mu$

$$
P_n = \begin{cases} 
\frac{1}{\lambda} \sum_{k=0}^{n} \left( \frac{\lambda}{\lambda - \mu} \right)^k P_0, & 0 \leq n \leq C \\
\frac{1}{C^n - \lambda, C} \left( \frac{\mu}{\lambda} \right)^n P_0, & C < n \leq C + N 
\end{cases}
$$

$$
D_q = \sum_{n=1}^{\infty} \frac{\lambda P(0)}{\sum_{n=1}^{\infty} \frac{C \mu P(0)}{C + n}}
$$

Equations (1) – (3) are used to evaluate the system performance using different input values.

5. Vertical Handover Performance Evaluation

In this section, we show input values for the model and simulation, and present the outcomes of the performance evaluation.

5.1 Input Values

The input values of both the model and simulation parameters are set as listed in Table 1. These values are used in the Markov model to calculate the values of parameters derived by the model. The simulation uses these values as initial values to run the simulation. Different runs of the simulation use different input values for analysis and comparison. For example, Figure 4. uses a range of values for the number of channels ($5 - 40$). Also, Figure 5. uses a range of values for the queue size ($1 - 100$). All results demonstrated in figures are found by the simulation unless stated.

It is worth mentioning that a Markov model generally states that when the arrival rate is increased such that it exceeds the service rate, the QoS of the system deteriorates. Nevertheless, the exact influence on QoS depends on factors such as the number of channels, the queue size and the average holding time (average call duration). The results that we present here attempt to show expected QoS for different inputs and the influence of each parameter on QoS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Input value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macrocell arrival rate</td>
<td>10 Call/sec</td>
</tr>
<tr>
<td>Microcell arrival rate</td>
<td>20 Call/sec</td>
</tr>
<tr>
<td>Service rate</td>
<td>15 Call/sec</td>
</tr>
<tr>
<td>Total number of channels</td>
<td>15</td>
</tr>
<tr>
<td>Queue size</td>
<td>50</td>
</tr>
</tbody>
</table>
5.2 Mean Queue Delay Measurements

![Graph showing mean queue delay](image1)

Figure 4: Mean queue delay having different number of channels obtained from the simulation and math model.

![Graph showing mean delay vs queue size](image2)

Figure 5: Mean queue delay (s) having different queue sizes obtained from the math model.
We evaluate the mean queue delay with respect to three parameters: number of channels, queue size and arrival rate.

5.2.1 **MEAN DELAY HAVING DIFFERENT NUMBER OF CHANNELS:**

When the number of channels increases, more requests are served immediately reducing the delay, as shown in Figure 4. The results also compare the simulation and the analytical model. As Figure 4 shows, the results are comparable, which proves the validity of the two approaches.

5.2.2 **MEAN DELAY HAVING DIFFERENT QUEUE SIZE:**

A longer queue implies storing more requests in the queue increasing the number of requests awaiting service. This results in increasing the average queue delay, as can be seen in Figure 5.

5.2.3 **MEAN DELAY HAVING DIFFERENT ARRIVAL RATES:**

The average queue delay increases with increasing the request arrival rate as more requests result in a longer queue. This continues until the queue is full, where any additional arrivals are dropped resulting in the saturation of the value of delay. Figure 6 shows the queue delay considering different arrival rates. It also shows the delay for queue sizes 10, 20 and 40. As can be seen in Figure 6, the delay for longer queues saturates at a higher value for the delay, consistent with findings shown in Figure 5.

5.3 **DROPPING PROBABILITY MEASUREMENTS**

![Figure 6](image6.png)

Figure 6 Mean queue delay having different arrival rates obtained by the math model

A longer queue implies storing more requests in the queue increasing the number of requests awaiting service. This results in increasing the average queue delay, as can be seen in Figure 5.

![Figure 7](image7.png)

Figure 7 Dropping probability having different number of channels when increasing the traffic (request/second)
The dropping probability is calculated using Equation (2) where $n$ is substituted with the system maximum capacity, as a handover request is dropped when the queue is full. Similar to the queue delay, the dropping probability is evaluated with respect to three parameters: the number of channels, the queue size and arrival rate.

5.3.1 DROPPING PROBABILITY HAVING DIFFERENT NUMBER OF CHANNELS:

When increasing the number of channels, more requests are served, reducing the number of rejected calls due to unavailable resources. The dropping probability decreases by increasing the number of channels as shows in Figure. 7.

5.3.2 DROPPING PROBABILITY HAVING DIFFERENT QUEUE SIZES:

The dropping probability is reduced by increasing the queue size, because the queue can handle a larger number of arrivals as can be seen in Figure. 8.

5.3.3 DROPPING PROBABILITY HAVING DIFFERENT ARRIVAL RATES:

The dropping probability increases with increasing the arrival rate. The queue reaches its maximum capacity faster and the remaining requests are dropped. Both Figure. 7 and Figure .8 show this increase.

5.4 MACROCELL-MICROCELL EVALUATION

To evaluate the influence of different-layer traffic on same-layer traffic, we consider three cases:

1. Case one: where macrocell and microcell traffic are the same.
2. Case two: macrocell arrival rate is 20 requests/second and microcell arrival rate is 10 requests/second.
3. Case three: macrocell and microcell request arrival rates are 10 and 20 requests/second, respectively.

For these cases, we evaluate the mean queue delay, the dropping probability and channel utilization. Channel utilization is found by calculating the fraction of time system channels are busy over total simulation time.
5.4.1 MACROCELL AND MICROCELL TRAFFIC IS SIMILAR

![Macrocell-Microcell mean queue delay](image1)

Figure 9 Macrocell-Microcell mean queue delay

![Macrocell-Microcell dropping probability](image2)

Figure 10 Macrocell-Microcell dropping probability

![Macrocell-Microcell channel utilization](image3)

Figure 11 Macrocell-Microcell channel utilization

In this case, traffic from both layers is the same. Macrocell and microcell requests experience the same delay as can be seen in Figure 9. The dropping probability and channel utilization are also the same for both layers as Figure 10 and Figure 11 confirm, respectively.
5.4.2 MACROCELL TRAFFIC IS DOUBLE THE MICROCELL TRAFFIC

Figure 12: Macrocell-Microcell mean queue delay

Figure 13: Macrocell-Microcell dropping probability

Figure 14: Macrocell-Microcell channel utilization
Here, the traffic of the macrocell is double that of the microcell. Observing Figure.12, the mean queue delay is similar for both layers as the queue treats requests from both layers equally. However, as requests generated at the macrocell are more, the dropping rate is more. Figure. 13 indicates that the dropping rate for macrocell traffic is double, consistent with the difference in traffic. Figure. 14 shows that the channel utilization is similar for both layers.

### 5.4.3 Macrocell Traffic Is Half The Microcell Traffic

In this scenario, the traffic of the macrocell is half the microcell traffic. The difference in mean queue delay is marginal as can be seen in Figure.15. The dropping rate follows the traffic ratio, as dropping rate for the macrocell is half of that of the microcell. Figure.16 shows this result. Similar to the previous case, the channel utilization is similar for both layers, as Figure.17 illustrates.

![Macrocell-Microcell mean queue delay](Figure.15)

![Macrocell-Microcell dropping probability](Figure.16)
6. CONCLUSIONS

This paper has investigated the performance of hard handovers in cellular mobile systems. It has evaluated a system consisting of a layer of macrocells overlapping a layer of microcells, where vertical hard handovers occur between the two layers. The evaluation is carried out using a simulation and a Markov model.

From the results it has been proven that when increasing the number of channels the mean queue delay and handover dropping rate are decreased. Yet this increase is restricted to the limitations in bandwidth. Increasing the queue size decreases the dropping rate at the expense of increasing the mean queue delay. Therefore an optimum queue size should be selected that results in an acceptable level of delay to maintain the guaranteed QoS and that is tolerated by the type of service. Finally, by increasing the arrival rate, the mean delay and the dropping rate increase. When traffic is expected to increase, resources should be provisioned to accommodate user traffic. The influence of traffic entering a cell from a different layer depends on the ratio of traffic between these layers. This statement is true when the two traffic types are treated with equal priority in the cell. In our future work, we plan to investigate applying different priorities to calls of each layer and draw results for different scenarios. We also plan to expand the work to cover soft handovers where more than one channel is allocated to the user during the handover process.

REFERENCES


Authors

Sahar Bashir Abbas received the B.Sc. degree in Computer Systems and Networks from Sudan University of Science and Technology, Khartoum, Sudan, in 2016. She is currently a network engineer. Her current research interests include modelling cloud-based storage networks and 4G networks.

Samah Fageer Ahmed received the B.Sc. degree in Computer Systems and Networks from Sudan University of Science and Technology, Khartoum, Sudan, in 2016. Her current research interests include modelling of inter system handovers and queuing networks.

Shimaa Hyder Makkì received the B.Sc. degree in Computer Systems and Networks from Sudan University of Science and Technology, Khartoum, Sudan, in 2016. Her current research interests include wireless communication systems and network performance modelling.

Niemah Izzeldin Osman received the B.Sc. degree (first class honours) in Computer Science from Sudan University of Science and Technology, Khartoum, Sudan, in 2002 and the M.Sc. degree (with distinction) in Mobile Computing from the University of Bradford, U.K., in 2006 and the Ph.D. degree in Communication Networks from the University of Leeds, U.K in 2015. She is currently an Assistant Professor at the Department of Computer Systems and Networks, Sudan University of Science and Technology, Sudan. Her current research interests include performance evaluation of 4G LTE networks, Internet of Things and QoE of video services.