INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 11(6):789-803, December 2016.

Analytical Modelling of a New Handover Algorithm for Improve Allocation of Resources in Highly Mobile Environments

Y. Kirsal

Yonal Kirsal*
Electrical and Electronic Engineering
European University of Lefke
North Cyprus, Mersin 10, Turkey
*Corresponding author: ykirsal@eul.edu.tr

Abstract: Wireless and mobile communication systems have evolved considerably in recent years. Seamless mobility is one of the main challenges facing mobile users in wireless and mobile systems. However, highly mobile users lead to a high number of handover failures and unnecessary handovers due to the limited resources and coverage limitations with a high mobile speed. The traditional handover models are unable to cope with high mobile users in such environments. This paper proposes, an intelligent handover decision approach to minimize the probability of handover failures and unnecessary handovers whilst maximizing the usage of resources in highly mobile environments. The proposed approach is based on modelling the system using a Markov chain to enhance the system's performance in terms of blocking probability, mean queue length and transmission delay. The results are compared with the traditional handover model. Simulation is also employed to validate the accuracy of the proposed model. Numerical results have shown that the proposed method outperforms the traditional algorithm over a wide range of handover failures and significantly reduced the number of such failures and unnecessary handovers. The results of this study show that quality if service (QoS) measures of such systems can be evaluated efficiently and accurately using the proposed analytical model. However, the performance results have also shown that it is still necessary to explore an effective model for operational spaces. In addition, the proposed model can also be adapted to various types of networks considering the high speed of the mobile user and the radius of the network.
Keywords: analytical modelling, mobility, handover decision algorithm, quality of service (QoS), highly mobile environments.

1 Introduction

With the rapid development and deployment of wireless technologies, next-generation wireless networks are expected to provide seamless mobility and ubiquitous access to the networks [1,2,4]. Researchers have focused on the improved quality of service (QoS) and performance evaluation of 4G/5G networks in highly mobile environments [6, 7, 9]. One of the main challenges for seamless mobility in next-generation wireless networks is the availability of the resources in the networks which allow mobile users to roam among heterogeneous environments [4, 9]. Mobility between wireless networks may lead to a high number of unnecessary handovers and handover failures when a mobile user is in highly mobile environments [10, 11]. The mobile user requires less time to across the coverage area of the network when the speed increases. Thus, the mobile user does not have enough time to acquire the network resources to do the handover. In addition, mobile users can join the queue by requesting resources from the system. However, mobile users will never get access to a channel for communication due to the user's velocity as well as the lack of resources. Hence, the queue will have unnecessary handovers by allowing such users to access the system. Unnecessary handover occurs when the mobile user's travelling time within the system is less than the handover process from the neighbouring networks to the system. Thus, the mobile user leaves the network coverage area before the handover process is executed [2, 9, 10]. This causes network connection breakdown [11] and interrupts the service [10]. Unnecessary handover is undesirable because it wastes network resources [7, 10]. On the other hand, if the user's call holding time is equal to or less than the total time of handover into and out of the system, then handover failure occurs [11]. In this situation, the mobile user does not transmit or receive any data packet to system; however, the mobile user might enter the system just after triggering the handover process. In other words, the system will be available after a short time when the mobile user triggers the handover process.

Many network characteristics (such as power consumption, received signal strength, and network conditions) affect the on handover process in wireless and mobile systems. Most existing decision algorithms have focused on the vertical handover process, such as optimization problems [12], a policy-enabled schemes [13, 14], and fuzzy logic [15]. On the other hand, in [16] the main focus is various mathematical models used in vertical handover decisions for heterogeneous networks. However, most previous works consider vertical handover decision algorithms. More recently, the literature has proposed various handover decision mechanisms [1–6] but none of the existing works consider user's the high mobility in the handover decision procedure. In [1] two-dimensional (2-D) Markov queuing models have been constructed to enhance multiple service requirements in the LTE network. The proposed models decrease the call blocking rate, especially for handover users. However, the mobility issues are not considered in [1]. The authors in [2] demonstrated their overall approach by describing the VANET Testbed in vehicular environments. The results obtained in [2] showed that it is necessary to consider a new handover model based on a probabilistic rather than traditional approach. In addition, proactive queueing approach for handover is also considered in [2], but the parameters (e.g., two channels in the systems) and scenario used for the analysis are rather simple in order to obtain realistic QoS results in highly mobile environments. In [4] a simple and robust two-step vertical handover decision algorithm is proposed for wireless and mobile networks. The new call blocking probability of the proposed model is modelled as $M/M/B_i/B_i$ based on the Erlang-B model. However, [5] showed that the Erlang-B model is not suitable for the handover schemes of highly mobile environments due to user's mobility in the system. In addition, as the reserved bandwidth may not be utilized effectively in low handover rates, the traditional reservation-based schemes are not efficient for future networks, especially in 4G vehicular networks [16]. In [17] the time before the vertical handover and the network dwell time are calculated and presented for any network topology. However, both parameters have not been used in previous studies to improve unnecessary handovers and handover failures.

In order to achieve seamless handover, when the mobile user has a high speed, it is important to predict the network availability, the time before handover and network dwell time with coverage boundaries. The Time before handover is the time after which the handover occurs, and the network dwell time is the time that the mobile user spends in the coverage area of the network. These two parameters are important in order to obtain the best time and place for the handover for mobile users. It is also possible to improve resource allocation in mobile and wireless systems by using these two parameters. Traditional handover models have been studied for wireless and mobile systems [1,3–5,8,9] but they lead to the degradation of the QoS due to the network's small coverage area and the velocity of the users, especially in highly mobile environments. Hence, the new handover approach is necessary for providing ubiquitous communication in next-generation systems. An analytical modelling and performance evaluation of mobile and wireless system using queueing theory has recently been performed [1–6, 8, 9]. Modelling limited resources and enhancing the QoS of the systems are the basis of the queuing phenomenon. Thus, the queuing theory can be used to model and analyse such problems [1, 7, 9]. Therefore, developing an analytical model of a new handover approach and resource allocation model for such systems would be the best option for obtaining more efficient QoS measurements.

This paper develops a new analytical model for handover and resource allocation in highly mobile environments based on the time the mobile user needs to acquire network resources for the handover. The main contribution of the paper can be summarized as follows:

- The QoS degradation of the traditional approach due to the handover failures and unnecessary handovers can be improved by the proposed algorithm considering call holding time, mobile user dwell time and time before handover.
- The performance improvement of high mobile users and management of the high mobile users within the cell and/or between neighbouring cells can be based on the acceptance factor.
- The results show that the proposed method gives better performance results than the traditional approach. However, a statistical model is still necessary to predict the degree of contention in highly mobile environments.

The main purpose of this paper is to develop a useful analytical model based on time before handover, call holding time and network dwell time by using a new decision algorithm to improve the QoS of real networks. The analysis done in this paper is based on modelling the system using a Markov chain to enhance the system performance in terms of blocking probability, mean queue length and transmission delay. This proposed model is applicable for most wireless communication systems. The rest of the paper is organised as follows: Section II presents the traditional handover approach. Section III describes the proposed handover approach. Section IV discusses the performance evaluation of the proposed algorithm with the traditional approach and simulation results. Finally, Section V provides the conclusions and future works.

2 The traditional approach

This section explains and represents the traditional handover approach for wireless and mobile environments. Figure 1 shows the handover process in such environments. The traditional handover approach introduces two types of threshold circles in the coverage area of the system [3, 8, 9, 17]. The handover threshold and the exit threshold circles are shown in Figure 1 for the traditional handover approach.

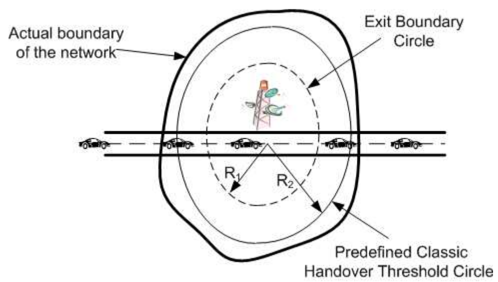

Figure 1: The traditional handover approach with threshold circles

Based on the traditional approach, the cell can be divided into different regions depending on the radius. The continuous circle with radius R_2 represents the handover threshold circle. In addition, the dotted circle with R_1 represents the exit threshold circle as shown in Figure 1. The exit threshold circle is the starting point for the the handover process in the traditional handover approach. In order to process successful mobility, mobile users have to finish the handover before reaching the classic handover threshold circle. If the mobile user is not handed over successfully before the circle, the mobile user will lose the connection. The traditional handover approach is currently being used in [1, 3–5, 8, 9] and [17] for wireless and mobile systems.

2.1 Traditional handover queueing model

In mobile and wireless communication networks, the service is provided by the base station (BS) and/or access point(AP) depending on the network. Mobile users communicate via radio links with BSs/APs [3, 5]. A single and arbitrary shape of cell is assumed. There are S channels that the system can provide for the service. In addition, the length of the queue is Q. The maximum number of calls allowed into the system is a combination of the number of users being served (S) and the number of users in the queue (Q). Hence, the maximum number of calls in the system is given by L where $L = S + Q$. The traditional handover queueing model is given in Figure 2 [3, 8, 9].

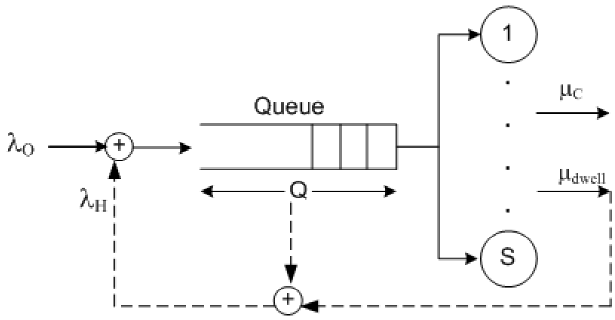

Figure 2: Traditional handover queueing model

The originating calls and handover calls are two kinds of arrival rates defined in the system, with mean arrival rates given as λ_O and λ_H, respectively. λ_O is newly generated calls in the system and λ_H is the mobile calls from one cell to another. If the channels are available and idle in the system, both call's arrival can be assigned to any channel. Otherwise, the incoming call request is added to the queue if the channels are busy [3,8,9]. In addition, if the queue is full, the incoming call is blocked. The channel requests in the system are served by first in first out (FIFO) rule. The inter-arrival times of the incoming call requests are assumed to follow an exponential distribution. λ is defined as the total arrival rate of calls in the cell, where $\lambda = \lambda_O + \lambda_H$. In traditional handover, the mobile users are moving at a velocity V, and there is a probability that it can also leave the network when being served due to the mobility. Moreover, a mobile user is placed in the queue waiting for the channel to be served. However, the mobile user can leave the system due to the mobility shown in Figure 2. A formula is given in [3] and [9] for λ_H. In traditional handover, T_C is call holding time in the system. An exponentially distributed T_C with a mean rate of μ_C is assumed. In addition, T_{dwell} is the dwell time, indicating the time that mobile users spend in the cell. This is also assumed to be exponentially distributed with a mean rate of μ_{dwell}. The equation 1 is used in the literature for the dwell time in wireless and mobile systems [3, 8, 9] for traditional handover queuing model. Thus, μ_{dwell} can be calculated and described as follows:

$$\mu_{dwell} = \frac{E[V] \cdot L}{\pi \cdot A} \tag{1}$$

where $E[V]$ is the average of the random variable, V is the speed of mobile users, L is the length of the perimeter of cell (a cell with an arbitrary shape is assumed), and A is the area of the cell [3, 8, 9]. The total channel holding time of a call is exponentially distributed with mean $1/\mu$ where, $\mu = \mu_C + \mu_{dwell}$. The state transition diagram of the traditional handover queueing model is shown in Figure 3. The states are defined as i (i=0,1,2,$\cdots$,S+Q) the number of calls in the system at time t.

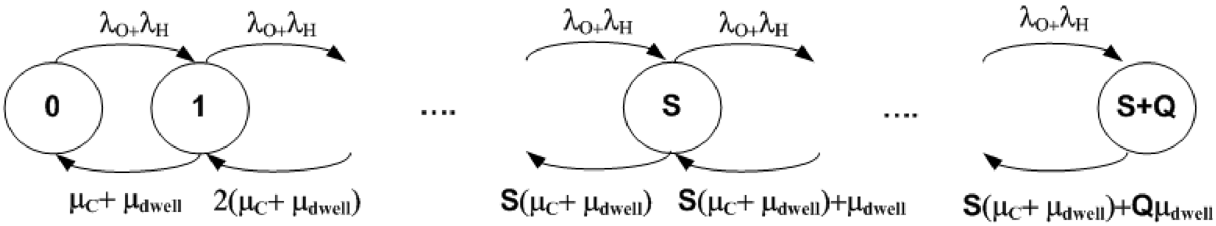

Figure 3: The state diagram of the traditional handover queueing model

ρ is the traffic intensity in the system, where ρ=λ/μ. Assuming a system in a steady state, the state probabilities, P_i's, can be obtained as in equation 2 [3, 8, 9].

$$P_i = \begin{cases} \frac{(\lambda_O+\lambda_H)^i}{i!} \cdot P_0 & 0 \leq i \leq S \\ \frac{\frac{(\lambda_O+\lambda_H)^S}{S!} \cdot (\lambda_O+\lambda_H)^{i-S} \cdot P_0}{\prod_{j=S+1}^{i} [S\mu + (j-S)\mu_{dwell}]} & S < i \leq S+Q \end{cases} \quad (2)$$

In equation 2, P_i is the probability that there are i calls in the system. P_0 can be defined as follows:

$$P_0 = \left[\sum_{i=0}^{S} \frac{(\lambda_O + \lambda_H)^i}{i!} + \sum_{i=S+1}^{S+Q} \frac{\frac{(\lambda_O+\lambda_H)^S}{S!} \cdot (\lambda_O + \lambda_H)^{i-S}}{\prod_{j=S+1}^{i} [S\mu + (j-S)\,\mu_{dwell}]} \right]^{-1} \quad (3)$$

Once all the steady state probabilities P_i are computed, the rest of the performance measures can be easily obtained. More information about the traditional handover queueing model can be found in [3, 8, 9].

3 The proposed approach

This section presents an abstract intelligent handover algorithm for high mobile environments applying queuing theory. The accurate knowledge of network availability, coverage boundaries (radius of the cell) and the velocity of mobile users are fundamental factors that play an important role in correct decision making during the handover. Hence, in the proposed model, the proposed algorithm determines the time that a mobile user needs before performing a handover. As mentioned in the previous sections, the proposed scheme is based on the current time $T_{current}$, network dwell time T_{dwell} and estimated time before handover $T_{estimated}$ of the mobile users. In order to reduce the number of unnecessary handovers and handover failures, the proposed algorithm determines the required time for the mobile user whether, admitting it into the system or performing a handover as shown in Figure 4. Assuming a mobile user is moving at a velocity V towards the system at $T_{current}$, the user can request a channel for communication. The user needs a channel at $T_{estimated}$ and releases the channel at ($T_{estimated} + T_{dwell}$). Based on the call holding time (T_C) of the calls, three possible conditions are proposed and analysed in Figure 4. Hence, the proposed approach can improve the resource allocation, especially in highly mobile environments.

- First condition: $(T_{estimated} + T_{dwell})_{n-1} < (T_{estimated})_n$
  If the channel needs time of the current user (n) is higher than the channel release time of the user being served (n-1), then the mobile user can enter the system seamlessly. This shows that the mobile user has enough time to get a place in the system to be served. In other words, unnecessary handovers $(T_{current} + T_{dwell}) < (T_{estimated})$ and partial handovers $T_C \leq (T_{dwell} + T_{estimated})$ do not occur.

- Second condition: $(T_{estimated})_{n-1} < (T_{estimated})_n$ and $(T_{estimated} + T_{dwell})_{n-1} < (T_{estimated} + T_{dwell})_n$
  For the second condition, the users are currently using the channels or waiting in the queue to be served. If the channels' release time of the users and/or waiting time in the queue (n-1) are higher than the channel release time of the current user (n), then the system will be partially busy by the time the current user reaches the system. This means that the current user might be admitted to the system after a short time. In other words, the system will be available soon for the service after the current user requests a channel. Hence, there is a partial contention $T_C \leq (T_{dwell} + T_{estimated})$ in the system.

- Third condition: $(T_{estimated})_{n-1} < (T_{estimated})_n$ and $(T_{estimated} + T_{dwell})_{n-1} > (T_{estimated} + T_{dwell})_n$
  If the channel release time of the users being served (n-1) is greater than the channel release time of the current user (n), then the system will be busy during the travel of the current user. Hence, the current user will never get access to the system. The channels and queue will no longer be available and the mobile user will be handed over to another network. Thus, the mobile user leaves the network coverage area before the handover process is executed $(T_{current} + T_{dwell}) < (T_{estimated})$ [11]. This causes a network connection breakdown [11] and interrupts the service [10].

In summary, in the event of the third condition, current mobile users will never join the system. The unnecessary handovers occur due to the high speed of the user as well as the radius of the network. Thus, the proposed algorithm passes the mobile user to the next available network via the acceptance factors. When the first condition is identified, the system (channels plus queue) can be used by the mobile user. In addition, when the second condition is identified and notified before the current user reaches the system, the contention can be signalled and the mobile user might be passed to other available networks nearby instead of waiting for the service. This approach should result in better network performance.

3.1 The proposed handover queuing model

In the proposed approach, the decision algorithm decides whether the mobile user will be admitted to the system based on the analysis described above section (see Figure 4). It is clearly seen that the proposed algorithm ensures that mobile users do not wait and leave the system unserved because of mobility. In other words, all mobile users will be allowed into the system depending on the analysis. Otherwise, the mobile user at a high speed towards the system will not have enough time to enter the system. Hence, the mobile users move at a high speed will be handed over to the another available network. Thus, mobile users do not wait long and leave the system without service. The proposed handover queueing model is shown in Figure 5.

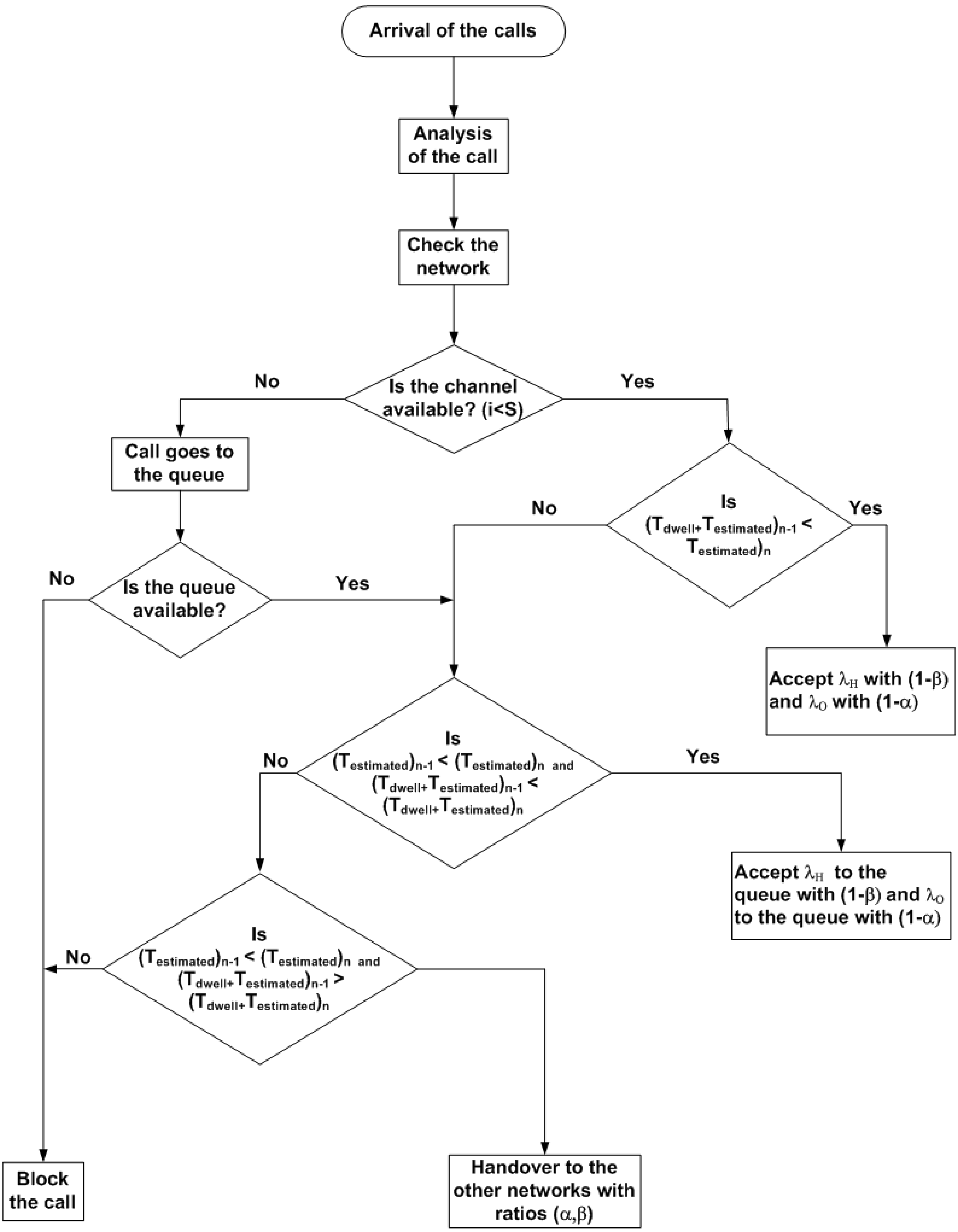

Figure 4: The proposed handover decision algorithm

The proposed handover queueing model (similar to the traditional handover) considers S number of channels and can allow i requests at time t as shown in Figure 5. The queueing capacity of the system is Q. The arriving requests may be sent from different users to the

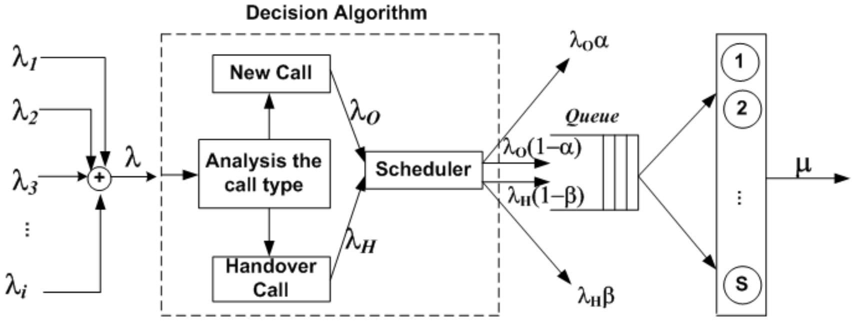

Figure 5: The proposed handover queueing model

system. Hence, the inter-arrival time of consecutive requests follows the Poisson process which can be distributed as an exponential distribution with arrival rate λ. According to [3], for a two-dimensional fluid model, the arrival rate of handover calls can be obtained as follows:

$$\lambda_H \approx \frac{\mu_{dwell}}{\mu_c}\lambda_O \tag{4}$$

The decision algorithm distinguishes the calls (λ_O/λ_H) and decides to send them into the system depending on the acceptance factors. α and β are the acceptance factors of the originating calls and handover calls, respectively. For the purpose of the proposed analytical model, α and β are taken as constant. It is assumed that the originating calls can join the system with an arrival rate of λ_O (1-α). Similarly, the handover calls can join the system with an arrival rate of λ_H (1-β). Hence, the total arrival rate is $\lambda = \lambda_O(1-\alpha)+\lambda_H$ (1-β). As the requests are rejected from entering the system, especially into the queue, the queue can be treated as a normal queuing system. Hence, the service rate is $\mu = \mu_C + \mu_{dwell}$.

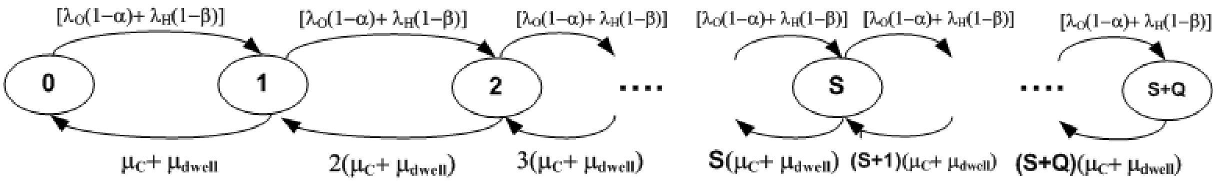

Figure 6: The state diagram of the proposed handover model

It is clearly seen that the M/M/C/K queuing model is fit for the proposed model for the performance evaluation. Thus, the proposed system can be illustrated by the given one-dimensional Markov chain as shown in Figure 6.

Let's define the states i (i=0,1,2,$\cdots$,S+Q) as the number of calls in the system at time t. The arrival rate can be taken as constant for all requests regardless of the number of users in the system. Hence, the arrival rate is the birth rate in the proposed model and can be obtained as [λ_O (1-α) + λ_H (1-β)]. In contrast, the rate of service completions in the proposed scheme depends on the number of calls in the system based on the analysis. If there are S or more requests in the system, then all S channels are busy. As each channel services users at the rate $\mu_C + \mu_{dwell}$, the combined service rate for the system is $S(\mu_C + \mu_{dwell})$. If there are fewer than S requests in the system, $i < S$, only i of the S channels are busy and the combined service rate for the system is $i(\mu_C + \mu_{dwell})$, as shown in Figure 6. Hence μ_i can be calculated as follows:

$$\mu_i = \begin{cases} i(\mu_C + \mu_{dwell}) & 0 \le i < S \\ S(\mu_C + \mu_{dwell}) & S \le i \le S+Q \end{cases} \tag{5}$$

Assuming the system is in a steady state, then using the well-known birth and death process, the steady state probabilities P_i can be obtained and are given in Equation 6:

$$P_i = \begin{cases} \frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{i!(\mu_C+\mu_{dwell})^i} \cdot P_0 & 0 \le n < S \\ \\ \frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{S^{i-S}S!(\mu_C+\mu_{dwell})^i} \cdot P_0 & S \le i \le S+Q \end{cases} \tag{6}$$

In order to find P_0 the normalization condition used since the probabilities must sum to 1, which gives:

$$P_0 = \left[\sum_{i=0}^{S-1} \frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{i!(\mu_C+\mu_{dwell})^i} + \sum_{i=S}^{S+Q} \frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{S^{i-S}S!(\mu_C+\mu_{dwell})^i}\right]^{-1} \tag{7}$$

The average number of packets in the system, MQL can then be calculated as $MQL = \sum_{i=0}^{S+Q} i \cdot P_i$ which gives:

$$MQL = \left[\sum_{i=0}^{S-1} i\frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{i!(\mu_C+\mu_{dwell})^i} + \sum_{i=S}^{S+Q} i\frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^i}{S^{i-S}S!(\mu_C+\mu_{dwell})^i}\right] \cdot P_0 \tag{8}$$

Similarly, the blocking probability P_B can be calculated as:

$$P_B = P(S+Q) = \frac{[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]^{S+Q}}{S^Q S!(\mu_C+\mu_{dwell})^{S+Q}} \cdot P_0 \tag{9}$$

In addition, the average queue length L_Q is:

$$L_Q = \sum_{i=S+1}^{S+Q} (i-S) \cdot P_i \tag{10}$$

Using Little's formula, the mean waiting time of channel requests in the queue can be calculated as follows:

$$E[T_w] = \frac{L_Q}{(1-P_B)[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]} \tag{11}$$

Hence, the average value of time of a call in a cell is:

$$E[T_s] = \frac{MQL}{(1-P_B)[\lambda_O(1-\alpha)+\lambda_H(1-\beta)]} \tag{12}$$

Let us define MQL_H as the mean number of handovers per user during its lifetime which can be calculated as:

$$MQL_H = \frac{MQL_H E[Tw]E[Tw]}{E[Ts]} \tag{13}$$

Therefore, the mean transmission delay of packet is calculated:

$$E[T_D] = MQL_H E[Tw] \tag{14}$$

4 Performance evaluation

This section presents numerical results in order to show the accuracy and effectiveness of the proposed analytical model of the new handover algorithm to improve resource allocation in highly mobile environments. In addition, the results obtained from the solution of the traditional and proposed approaches are validated by using discrete event simulation (DES). The simulation tool is mainly used for the validation of both the traditional and proposed models. As the simulation simulates the actual scenario rather than the Markov models presented in this paper, it can also be used for the performance evaluation of such systems. The DES developed considers the stochastic processes for all types of mobile users' arrivals and departures. Mobile users' arrivals and departures occur one at a time in a random, discrete, event-triggered fashion when an arrival enters the system and service is completed, respectively. In addition, the users waiting in the system are served based on first come first serve (FCFS) basis in the order of their arrival. The channel (and/or channels) becomes idle or remains busy with requests stored in the queue when the service event is completed. While a particular event is handled, the next event is generated. The results obtained from the simulation runs are within the 5% confidence interval with a 95% confidence level [9]. The simulation model was adopted for the scenario considered and implemented in C++ language. In order to validate the proposed analytical model, the results obtained from the analytical model and the simulation results for different performance measures are presented and compared. The numerical study focuses on MQL, P_B and transmission delay of the proposed models. The mean arrival and mean service rates are mainly application dependent. The assumptions in [3, 5, 7, 8] and [9] are employed in this paper as well for consistency, unless stated otherwise.

4.1 Key parameters

The system parameters used are mainly taken from [3, 5, 7, 8] and [9] based on the relevant literature [1, 2, 4, 6, 10–13, 16]. The system has a fixed number of identical channels: S=16. Q is the queuing capacity which represents the number of packets waiting for service. It is assumed that the moving direction of the mobile users can be detected by the BS/AP using a control channel. In addition, a mixed traffic pattern is also assumed, as in [2] where on average a minimum of 2 slots are 0.5 ms. Hence, the rates are translated into packet per second in order to use consistent values. The service rate of the mobile users μ_{dwell} is calculated using Equation 1. The requests are handed over or rejected from entering the system due to the proposed analysis; thus, the arrival rate is $\lambda = \lambda_O(1-\alpha) + \lambda_H(1-\beta)$. However, in this paper α is taken as 0.01 because λ_H passing through in a unit time with a high speed is larger than λ_O. In other words, λ_O calls are assumed to be allocated by the system as they are newly generated in the system. The analysis of α could be explored in future work.

4.2 Results

The Figures 7 and 8 show MQL and P_B results, respectively, as a function of the originating calls λ_O in the system. The parameters are S=16, Q=50, $E[V]$=40m/s (144km/hr), R=1000m, $E[T_c]$=120 packets/sec, and α=0.01 and the λ_O rate per user varies from 0.01 packets per second.

The figures clearly show that the proposed approach works far better than the traditional approach. In the traditional approach, due to the high mobile users, most users will leave the system without being served. In addition, the handover calls from the neighbour cells will request channel allocation at the same time, especially for heavy traffic loads (e.g., λ_O=0.08). This causes an increase in MQL as well as the P_B of the system. Thus, β has an impact on the

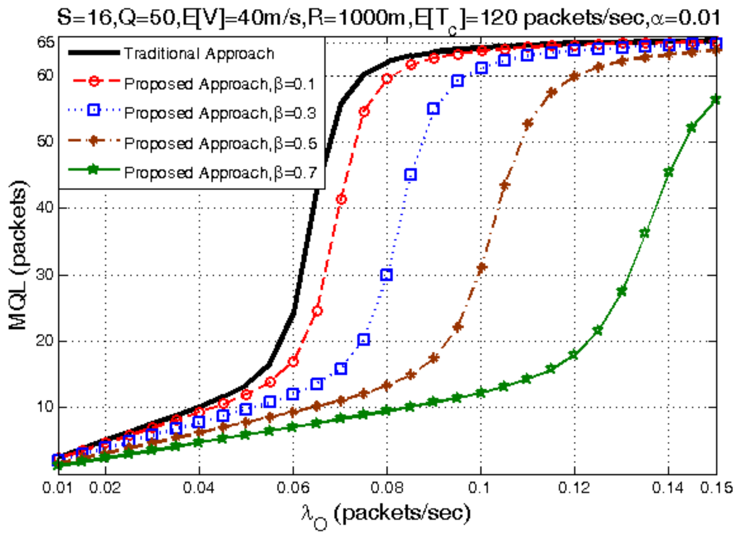

Figure 7: Mean queue length results as a function of originating calls λ_O with different β values

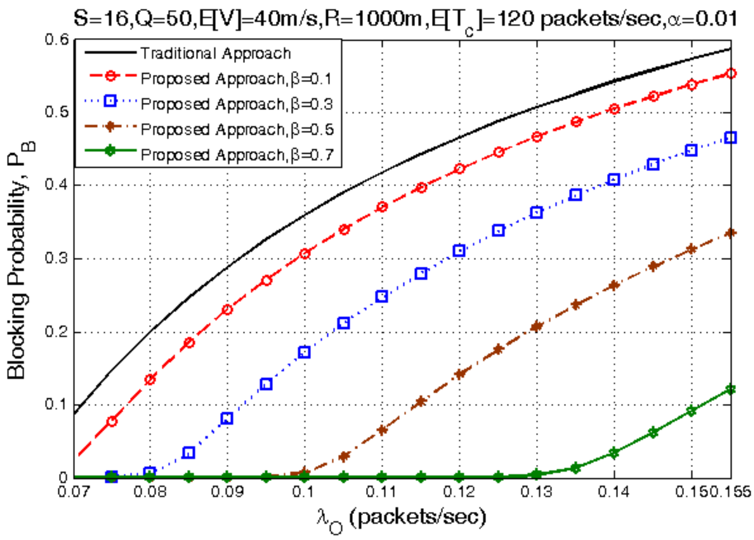

Figure 8: Blocking Probability results as a function of originating calls λ_O with different β values

system. It is clear from the figures that β significantly affects the system performance. Hence, β is an important parameter for the handover management in highly mobile environments.

Figure 9 shows transmission delay as a function of the originating calls λ_O with different β values. In wireless and mobile networks, transmission delay is another important QoS parameter criterion. It can be clearly seen that transmission delay increases rapidly for the traditional handover when λ_O increases due to the number of unnecessary handovers allowed in the system. Such handovers will leave the system without being served. The system is then busy with unnecessary handovers and the transmission delay increases. It can be observed from the graphs in Figures 7, 8 and 9 that the proposed approach gives better QoS results in terms of mean calls in the system, blocking probability and transmission delay, respectively, when β increases. This means that, according to the acceptance factor, highly mobile users are handed over to the neighbour cell and/or served without wasting the network resources.

Table 1 illustrates blocking probability results as a function of queue size. Parameters used

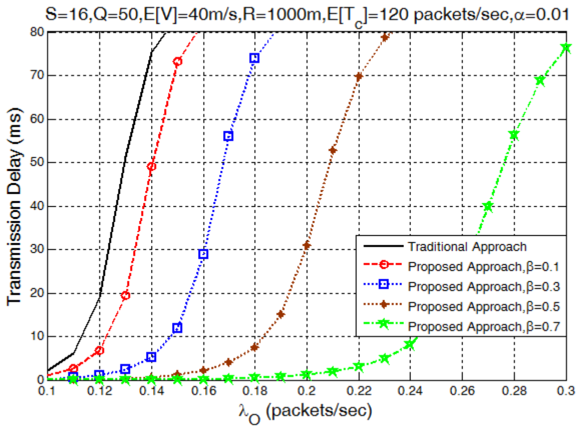

Figure 9: Transmission delay as a function of originating calls λ_O with different β values

Table 1: Blocking Probability results as a function of queue size (Q)

Q	Traditional Approach	Proposed Approach, $\beta = 0.2$	Proposed Approach, $\beta = 0.4$	Proposed Approach, $\beta = 0.6$	Proposed Approach, $\beta = 0.8$
30	0.089398	0.005699	1.57E-05	3.01E-09	1.35E-14
50	0.0863125	0.001211	6.89E-08	9.82E-14	6.58E-22
70	0.0858134	0.000269	3.03E-10	3.20E-18	3.20E-29
90	0.0857308	0.0000601	1.33E-12	1.04E-22	1.56E-36

for Figures 7, 8 and 9 are used for the results presented in Table 1 as well. The parameters are as follows: S=16, λ_O=0.1, $E[V]$=40m/s, R=1000m, $E[T_c]$=120 packets/secs and α=0.01. The blocking probability decreases slightly in the traditional handover because (especially for a loaded system) highly mobile users make the system busy due to the unnecessary handover as well as the handover failures. However, this is not the case when the proposed algorithm is employed. The blocking probability decreases rapidly when Q increases. This means that the proposed approach can handle unnecessary handovers and handover failures. In other words, the proposed approach gives better resource usage than the traditional approach.

On the other hand, MQL results as a function of originating calls for low mobile users are given in Figure 10. The results show that the proposed model performs better than the traditional approach when the system utilisation ($U = \lambda/S\mu$) is less than 0.72. However, as the velocity decreases, the traditional approach outperforms the proposed approach in some situations, especially for a heavy-loaded system (e.g., U=0.88). This is mainly because at such low velocity no one leaves the system due to the mobility. Then, large MQL results are experienced in the proposed system when $\beta = 0.1$ and 0.3. However, the proposed model gives better results when higher values of β are considered (i.e., $\beta = 0.7$). In addition, even at low velocity, most of the mobile users leave the queue without being served in the traditional approach.

The numerical results obtained from the proposed model are also validated by the simulation in Table 2 and Figure 11. The parameters used in Table 2 and Figure 11 are the same parameters used in Figures 7 and 8. Table 2 shows the P_B results of the traditional and proposed approaches with different β. It is obvious in Table 2 that numerical results obtained from the proposed model show agreement with the results obtained from the simulation as the discrepancies are less than 5%. The numerical results show the effectiveness of the proposed model. The MQL results for both approaches as a function of λ_O are shown in Figure 11 and validated by simulations. The results of the proposed analytical approach and simulation results show good agreement. The maximum discrepancy between the analytical model and simulation is 3.42% which is well within

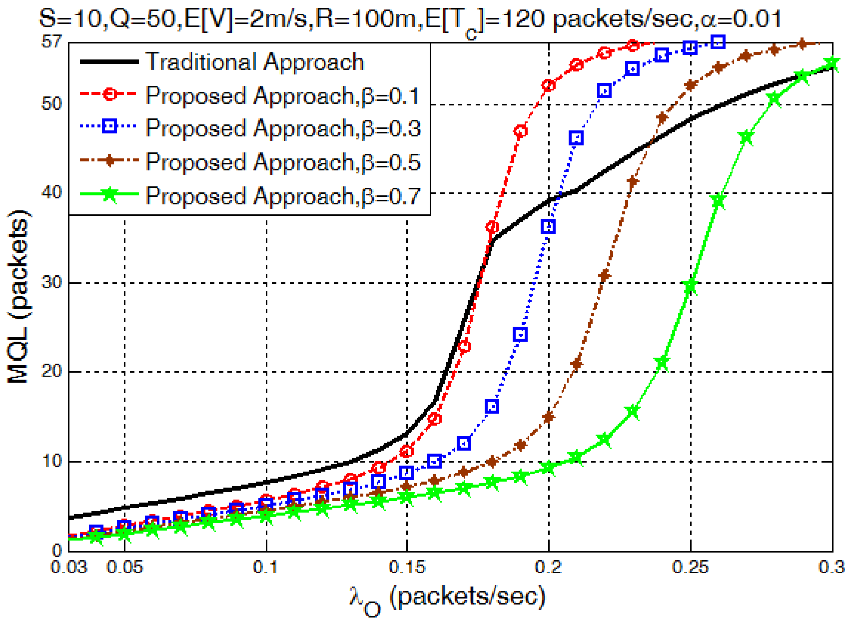

Figure 10: Mean queue length results as a function of originating calls λ_O with different β values for low mobile users

the 5% confidence interval of the simulation.

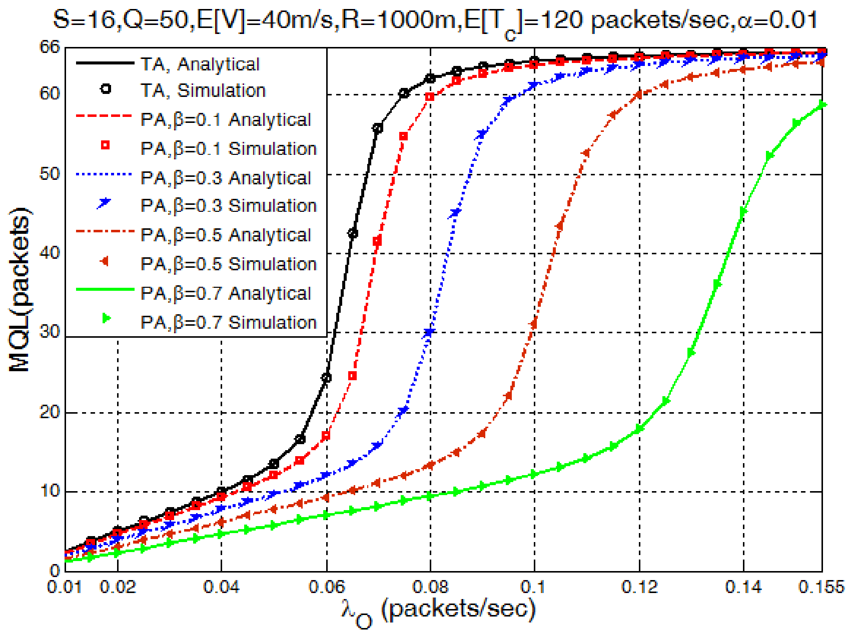

Figure 11: The analytical and simulation MQL results of both traditional approach (TA) and proposed approach (PA) as a function of originating calls λ_O with different β values

Table 2: Validation of P_B results as a function of λ_O for both traditional and proposed approaches. (D is Discrepancy).

	P_B,Traditional Approach			P_B,Proposed Approach β= 0.3			P_B,Proposed Approach β =0.5		
λ_O	Analytical	Simulation	D(%)	Analytical	Simulation	D(%)	Analytical	Simulation	D(%)
0.085	0.2471	0.2490	0.78	0.0335	0.0340	1.40	1.20E-12	1.20E-12	0.04
0.09	0.2889	0.2900	0.38	0.0805	0.0810	0.57	2.89E-11	2.89E-11	0.02
0.095	0.3263	0.3180	2.62	0.1282	0.1240	3.42	5.63E-10	5.62E-10	0.05
0.1	0.3600	0.3610	0.28	0.1718	0.1750	1.85	9.06E-09	9.05E-09	0.07
0.105	0.3905	0.3900	0.12	0.2112	0.2130	0.84	1.22E-07	1.22E-07	0.03
0.0.11	0.4182	0.4120	1.50	0.2471	0.2400	2.94	1.39E-06	1.39E-06	0.22
0.115	0.4435	0.4430	0.11	0.2798	0.2790	0.28	1.34E-05	1.35E-05	0.51
0.12	0.4667	0.4660	0.14	0.3098	0.3090	0.26	1.09E-04	1.10E-04	0.91
0.125	0.4880	0.4800	1.67	0.3374	0.3370	0.12	7.23E-04	7.30E-04	0.93
0.13	0.5077	0.5000	1.54	0.3629	0.3620	0.25	3.76E-03	7.70E-03	1.63
0.135	0.5259	0.5250	0.18	0.3865	0.3860	0.13	1.38E-02	1.3E-02	0.23
0.14	0.5429	0.5400	0.53	0.4084	0.4080	0.10	3.433E-02	3.43E-02	0.10
0.145	0.5586	0.5500	1.57	0.4288	0.4300	0.28	6.233E-02	6.23E-02	0.07
0.15	0.5733	0.5700	0.58	0.4478	0.4490	0.26	9.21E-02	9.39E-02	1.96
0.155	0.5871	0.5800	1.22	0.4657	0.4656	0.01	1.21E-01	1.22E-01	0.81

Conclusions and future work

This paper proposed a new analytical modelling approach and QoS management for handovers based on a new handover admission control mechanism in highly mobile environments. The analysis of the handover is an important issue in order to achieve better performance, especially in highly mobile environments. The proposed handover admission control mechanism is useful for achieving better performance in such systems. It offers the perspective of considering the current time $T_{current}$, network dwell time T_{dwell} and estimated time before handover $T_{estimated}$ of the mobile users. The system is modelled as an open queuing network using a Markov chain with continuous time to determine the state probabilities. Based on the proposed approach developed in this paper, computer simulations are also used to assess the accuracy for the proposed model. The proposed model can be used to analyse QoS measures such as MQL, P_B and transmission delay. The presented examples were kept simple for performance evaluation due to the introductory nature of the proposed model for highly mobile environments. The proposed method successfully reduced the number of handover failures and unnecessary handovers to the system by using the proposed algorithm compared to the traditional approach for highly mobile users. It minimizes the number of handover failures and unnecessary handovers to the system by enhancing the usage of the resources. With this approach, resource allocation can be improved in such systems with highly mobile environments. However, there are still specific operational aspects that need to be explored where the proposed approach can be applied to get the best effect. In addition, considering the availability, modelling the proposed model could be considered for future work.

Bibliography

[1] Chen, Y.; Yang, S.; Xu, S.; Xue, P.; Zhou, X.; (2012); Queuing Theory Based Handover Resource Self-Management in LTE Networks, *International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*:1-4.

[2] Ghosh, A.; Paranthaman, V.; Mapp, G.; Gemikonakli, O.; Loo, J. (2015); Enabling Seamless V2I Communications: Towards Developing Cooperative Automotive Applications in VANET Systems, *IEEE Communications Magazine, Special Issue on Towards Autonomous Driving: Advances in V2X Connectivity*, 53(12):80-86.

[3] Zeng, Q.A.; Agrawal, D. P., (2001); Modeling of handoffs and performance analysis of wireless data networks, *International Conference on Parallel Processing Workshops*, 491-496.

[4] He, D. et al. (2010); A simple and robust vertical handoff algorithm for heterogeneous wireless mobile networks, *Wireless Personal Communication*, 59(2):361-373.

[5] Trivedi, K.S.; Dharmaraja, S.; Ma, X.; (2002); Analytic modelling of handoffs in wireless cellular networks, *Information Sciences*, 148:155-166.

[6] Rejeba, S. B.; Nasser, N.; Tabbane, S.; (2014) A novel resource allocation scheme for LTE network in the presence of mobility, *Journal of Network and Computer Applications*, Elsevier, 46:352-361.

[7] Halabian, H.; Rengaraju, P.; Lung, C.H.; Lambadaris, I.; (2015), A reservation-based call admission control scheme and system modeling in 4G vehicular networks, *EURASIP Journal on Wireless Communications and Networking*, 1-12.

[8] Kirsal-Ever, Y.; Kirsal Y.; Ever, E.; Gemikonakli,O.; (2015); Analytical Modelling and Performability Evaluation of Multi-Channel WLANs with Global Failures, *International Journal of Computers Communications and Control*, 10:551-566.

[9] Kirsal Y.; Ever, E.; Kocyigit, A.; Gemikonakli,O.; Mapp, G.; (2015); Modelling and analysis of vertical handover in highly mobile environments, *The Journal of Supercomputing*, 71(12):4352-4380.

[10] Xiaohuan, Y., Mani, N., Sekercioglu, Y.A. (2008); A traveling distance prediction based method to minimize unnecessary handovers from cellular networks to WLANs, *IEEE Communications Letters*, 12(1):14-16.

[11] Kyoung, S.L., Ae-Soon, P. (2014); Reduction of handover failure for small cells in heterogeneous networks, *Intl Conf. on Information and Communication Technology Convergence (ICTC)*: 707-708.

[12] Zhu, F.; MacNair, J.; (2004); Optimizations for Vertical Handoff Decision Algorithms, *IEEE Wireless Communications and Networking Conference*, 2:867-872.

[13] Zhu, F.; McNair, J.; (2006); Multiservice vertical handoff decision algorithms, *EURASIP Journal on wireless communications and networking*, 2; 1-13.

[14] Stevens-Navarro, E.; Lin, Y.; Wong, V.W.S.; (2008); An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks, *IEEE Transactions on Vehicular Technology*, 57(2): 1243-1254.

[15] Ismail, A.; Byeong-hee, R.; (2011) Adaptive handovers in heterogeneous networks using fuzzy MADM, *Intl Conf. on Mobile IT-Convergence*, 99-104.

[16] Yan, X.; Sekercioglu, A. Y.; Narayanan, S.; (2010) A survey of vertical handover decision algorithms in fourth generation heterogeneous wireless networks, *Computer Networks*, 54(11):1848-1863.

[17] Mapp. G. et al. (2009); Exploring Efficient Imperative Handover Mechanisms for Heterogeneous Networks, *International Symposium on Emerging Ubiquitous and Pervasive Systems*, 286-291.