INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(5), 837-852, October 2018.

Exploring Analytical Models for Proactive Resource Management in Highly Mobile Environments

Y. Kirsal, V.V. Paranthaman, G. Mapp

Yonal Kirsal*
European University of Lefke
Faculty of Engineering
Department of Electronics and Communication Engineering
Lefke, North Cyprus
TR-10 Mersin, Turkey
*Corresponding author: ykirsal@eul.edu.tr

Vishnu Vardhan Paranthaman, Glenford Mapp
School of Science and Technology
Middlesex University
London, NW4 4BT
v.paranthaman@mdx.ac.uk, g.mapp@mdx.ac.uk

Abstract: In order to provide ubiquitous communication, seamless connectivity is now required in all environments including highly mobile networks. By using vertical handover techniques it is possible to provide uninterrupted communication as connections are dynamically switched between wireless networks as users move around. However, in a highly mobile environment, traditional reactive approaches to handover are inadequate. Therefore, proactive handover techniques, in which mobile nodes attempt to determine the best time and place to handover to local networks, are actively being investigated in the context of next generation mobile networks. Using this approach, it is possible to enhance channel allocation and resource management by using probabilistic mechanisms; because, it is possible to explicitly detect contention for resources. This paper presents a proactive approach for resource allocation in highly mobile networks and analysed the user contention for common resources such as radio channels in highly mobile wireless networks. The proposed approach uses an analytical modelling approach to model the contention and results are obtained showing enhanced system performance. Based on these results an operational space has been explored and are shown to be useful for emerging future networks such as 5G by allowing base stations to calculate the probability of contention based on the demand for network resources. This study indicates that the proactive model enhances handover and resource allocation for highly mobile networks. This paper analysed the effects of β and α, in effect how these parameters affect the proactive resource allocation requests in the contention queue has been modelled for any given scenario from the conference paper "Exploring analytical models to maintain quality-of-service for resource management using a proactive approach in highly mobile environments" [9] (doi:10.1109/ICCCC.2018.8390456).[a]

Keywords: analytical modelling, proactive resource management, contention, time before handover, time to handover, highly mobile environments.

[a]Reprinted (partial) and extended, with permission based on License Number 4386390665938 ©[2018] IEEE 7th International Conference on Computers Communications and Control (ICCCC)

1 Introduction

With the rapid development of mobile communication technologies such as WiFi, femto-cells, Long-Term Evolution Advanced (LTE-A), Vehicular Ad-Hoc Network (VANET), integration

Copyright ©2018 CC BY-NC

of various other wireless networks known as heterogeneous networking (HetNet) has become necessary to provide the mobile node (MN) with ubiquitous communication. For instance, where, a MN can connect to a nearby LTE, the calls rejected by LTE networks due to lack of radio access can overflow to overlaying WiFi networks, thus reducing the call blocking probability and improving bandwidth utilization in cellular overlay networks. However, this may result in frequent forced handover in those areas covered by small cells and leads to extra signalling overheads [8]. In these environments, traditional handover techniques, which depend on a reactive approach, have been found to be inadequate because of high speeds as resources must be quickly allocated and deallocated as the mobile user moves around. Hence, good resource management must be considered as a key enabling functionality to allow seamless connectivity in highly mobile heterogeneous environments.

The advent of latest wireless technologies, high data rate multimedia services and heterogeneous user equipments, necessitate the development of efficient techniques for resource management in highly mobile heterogeneous environments [15]. In addition, the high mobility between HetNets may lead to a high number of unnecessary handovers as well as handover failures due to the user's velocity [6]. Unnecessary handover occurs when the mobile user's dwell time within the network coverage 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 [20]. This causes network connection interruption, thus Quality of Service (QoS) of the system degrades. On the other hand, handover failures may lead to data loss, long delays, and even communication interruptions, which are not tolerable for safety life critical applications such as accident prevention, emergency and disaster management as required in Intelligent Transport Systems (ITS) [1].

Proactive behaviour by systems refers to anticipatory, self-initiated and change-oriented behaviour in situations. Proactive behaviour involves acting in advance of a future situation, rather than just reacting. It means taking control and making things happen rather than just adjusting to a situation or waiting for something to happen. Therefore, this proactive approach helps any system to perform better than a reactive approach [17]. It is very important to provide a seamless service to all the users in a network and therefore, these users and their contention for resources have to be looked in detail. In this context, client-based handover can be more efficient, since the metrics for the handover decision mechanism can be monitored by the user equipment (UE) from its wireless interfaces and can be used to decide to trigger the handover. In addition to the need for new handover decision mechanisms, there is also a need for better resource reservation mechanisms due to varying traffic characteristics, QoS application requirements and wireless channel conditions at the access point (AP). Efficient resource management is needed to optimize the performance of a wireless network because only a limited number of simultaneous calls can be hosted by a wireless cell. Therefore, incoming handover calls and new calls should not compromise the quality of the ongoing calls in the cell. Traditionally, user contention has been used to analyse the need for specific resources such as radio channels by mobile users. However, it is possible to use this contention to proactively manage these resources. Implementing the appropriate pre-planned resource allocation when the system state changes are known as proactive resource management. It provides the ability to specify end-to-end QoS quickly and to determine whether the specified QoS can be provided in advance of when these resources will be required [3].

Y-Comm is an architecture that has been designed to build future mobile networks by integrating communications, mobility, QoS and security [13]. The researchers of Y-Comm have explored new parameteres such as time before handover (TBH) and network dwell time (NDT) as shown in Figure 1 [18]. These two parameters was introduced in [16] and are used to provide a new proactive approach for handover and resource allocation in heterogeneous environments.

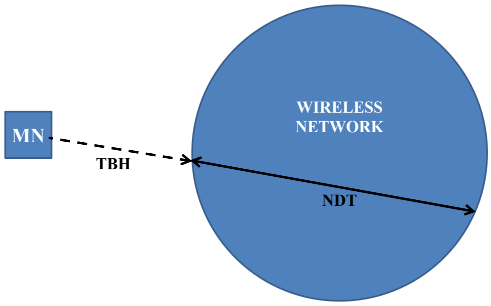


Figure 1: Illustrating time before handover (TBH) and network dwell time (NDT)

Thus, in this paper, possible contention probabilities are considered and operational areas are explored for certain scenarios using an analytical modelling approach. No contention, partial contention and full contention are three possible contentions that are considered based on two key parameters i.e., time to get resource ($\mathbb{T}$) and resource hold time ($\mathbb{N}$). $\mathbb{T}$ is the time when actual resource requested is available for use i.e., even after entering the network's coverage range with a successful soft handover, the resource required by the MN might not be available, for example, other users might be holding the resource. $\mathbb{N}$ is the resource usage time or when actual exchange of data is taking place. Thus, in order to achieve seamless handover in highly mobile environments, it is important to predict the resource availability considering $\mathbb{T}$ and $\mathbb{N}$.

This paper extends the conference paper [9]. The key additions of this journal version are as follows. It adds more on methods to formulate β and α more accurately for any given scenarios. A proactive approach used in [9] is analysed for resource allocation to formulate β and α by investigating all contention probabilities for channel access to a wireless network, results are obtained and verified using analytical models. The proposed model and results presented improve the capacity and services delivered by mobile networks. The rest of the paper is organized as follows: Section 2 includes and describes more related works. Section 3 gives details of resource allocation and contention in highly mobile environments. Section 4 gives a detailed description of our proposed model with the derived Markov queuing model for the new proactive approach and Section 5 shows the results and effects of β and α. Section 6 concludes and gives the future direction of the work.

2 Related work

Bandwidth is an extremely valuable and scarce resource in mobile networks. Therefore, efficient mobility-aware bandwidth reservation is necessary to support multimedia applications (e.g., video streaming) that require QoS. Several research efforts were carried out looking at routing, security and applications for highly mobile environment but very few addressed handover and resource allocation issues.Recent works in [2] and [11] clearly show that researchers are interested in proactive handover or predictive handover mechanisms, however these efforts considered parameters like user preferences, user location and application requirements.

Several studies of the vertical handover procedure, mobility management and common radio resource management schemes in heterogeneous environments have been reported in the literature [5, 21, 24]. These studies show that there are several dynamic factors that must be considered in vertical handover decisions for effective network usage including policies to determine whether or not to handover should occur as well as mechanisms to determine the best network to handover.

The work in [4] has proposed a model for enhancing the modelling of vertical handover in integrated cellular and wireless local area network environments. In addition, in [10] an intelligent handover decision approach is proposed to minimize the handover failures and unnecessary handovers whilst maximizing the usage of resources in highly mobile environments. On the other hand, QoS and Carrier to Interference-and-Noise Ratio (CINR) are some common metrics that must also be considered in this context. Any proposed solution must also be scalable because in the future the MN may have the ability to handover to hundreds of possible target networks. In this context, client-based handover can be more efficient, since the metrics for the handover decision mechanism can be monitored by the UE from its wireless interfaces and can be used to decide to trigger the handover [23]. Knowing the velocity and current position of an MN could help to estimate where the MN is heading, thus the next position of MN where handover might be performed can be predicted.

Proactive handover in which the MN actively attempts to decide when and where to handover has been shown to be an efficient handover policy mechanism to minimize packet loss and service disruption as an impending handover can be signalled to the higher layers of the network protocol stack [12]. A mobility prediction scheme for MNs was proposed in [14]. Probability and Dempster-Shafer processes were applied to predict the likelihood of the next destination based on the user's habits (e.g., frequently visited locations). In addition, second-order Markov chain process was applied at each road junction for predicting the likelihood of the next road segment transition, given the direction to the destination and the path from the trip origin to that specific road junction. A proactive unnecessary handover avoidance scheme was proposed in [22] for the LTE-A small cells. Unnecessary handover was avoided by calculating the probability of the active time during the dwell time in one cell and comparing the same to find if pre-defined threshold exceeds certain value. Here, a model to estimate the dwell time in the small cell was not considered. In [7], Fernandez *et al.* highlighted the need for a decision mechanism for choosing an appropriate point of attachment for high mobility nodes.

All studies above suggested resource management as well as handover guidelines in order to achieve seamless communication in high mobility environments. However, in this paper the work in [9] has been extended and analysed the request not being affected and staying in queue considering all possible contention probabilities in highly mobile environments based on the time the mobile user needs to acquire network resources before the handover.

3 Resource allocation and contention in mobile networks

In the era of increasing connectivity, wireless networks are gaining importance in several applications. One of their obvious benefits is the support of mobile users. In complex network infrastructures consisting of many APs, seamless connectivity becomes an important aspect because many applications rely on real-time communication and therefore need seamless handover between APs. If there is a decrease in quality the connection below a minimum threshold then the connection is lost, and the client must scan the wireless channels and look for a new AP in order to establish a new connection. The scanning procedure typically takes a long time and thus prohibits seamless connectivity [6]. Most of the analysis on handover and resource allocation was service oriented and it is important to look into the impact on individual users based on their mobility and dwell times in a network.

In mobile environments it is also necessary to understand why contention should be considered and this can be done by looking at classical approach to analyse the performance of mobile networks. Classical handover occurs when a MN changes its points of attachments (PoA) from the current wireless network to another network using a reactive approach i.e., the handover is initiated only after the MN is within the network coverage of the next wireless network. Classical handover uses a reactive approach i.e., when the MN is moving away from the area covered by one cell and entering the area covered by another cell the call is transferred to the second cell in order to avoid call termination when the MN gets outside the range of the first cell. In this approach in order to start the handover, the MN should be in the coverage range of the second cell and has to exchange the relative information to start the handover and complete the handover before exiting the first cell.

There are a number of parameters that need to be known to determine whether a handover is required. The signal strength of the base station (BS) with which communication is being made, along with the signal strengths of the surrounding stations. Additionally the availability of channels also needs to be known. The MN is obviously best suited to monitor the strength of the BSs at its location, but only the cellular network knows the status of channel availability and the network makes the decision about when the handover is to take place and to which channel of which cell [24].

Accordingly, the MN continually monitors the signal strengths of the BSs it can hear, including the one it is currently using, and it feeds this information back to the BS. When the strength of the signal from the BS that the MN is using starts to fall to a level where action needs to be taken then the cellular network looks at the reported strength of the signals from other cells reported by the MN. It then checks for channel availability, and if one is available it informs this new cell to reserve a channel for the incoming MN. When ready, the current BS passes the relevant information for the new channel to the MN, which then makes the change. Once there the MN sends a message on the new channel to inform that it is now within the coverage of the new network. If this message is successfully sent and received then the network shuts down communication with the MN on the old channel, freeing it up for other users, and all communication takes place on the new channel.

Under some circumstances such as when one base transceiver station is nearing its capacity, the network may decide to hand some MNs off to another base transceiver station they are receiving that has more capacity, and in this way reduces the load on very busy the base transceiver station. Hence, access can be opened to the maximum number of users. In fact channel usage and capacity are very important factors in the design of a cellular network.

In [10] a service oriented approach for MNs is presented, considering that the services will be resumed as soon as the MN moves to a different network. But the work did not focus on the resources available at the AP or BS and the effects of mobility in acquiring this resource. Understanding this effect is very important as the MNs would be waiting to acquire a channel and might move out of the current network due to mobility without service. The problem with the classical handover approach is that the AP/BS does not know in advance the network requirements of MNs heading towards it. Due to this the MN even after entering the network coverage, it will have to wait until the resource becomes free. In a highly mobile environments, MNs will have less time to spend in a network coverage therefore, the AP/BS has to anticipate the network conditions much before based on the MNs about to reach its network in order to have an effective resource utilization.

3.1 Network coverage parameters for mobile networks

In this work we are further exploring and redefining the communication range segments [15] as shown in Figure 2 which can be put into effective use for achieving both proactive handover and resource allocation for a highly mobile environment.

Figure 2 shows a more advanced scenario in which three consecutive overlapping wireless networks are segmented based on various key time variables which can be used to enhance handover and resource allocation. Time before handover (Υ) is the time after which the handover process should start and Time to handover ($\hbar$) is the time before which the handover to next coverage range has to be completed, if not it will result in a hard handover. Network Dwell

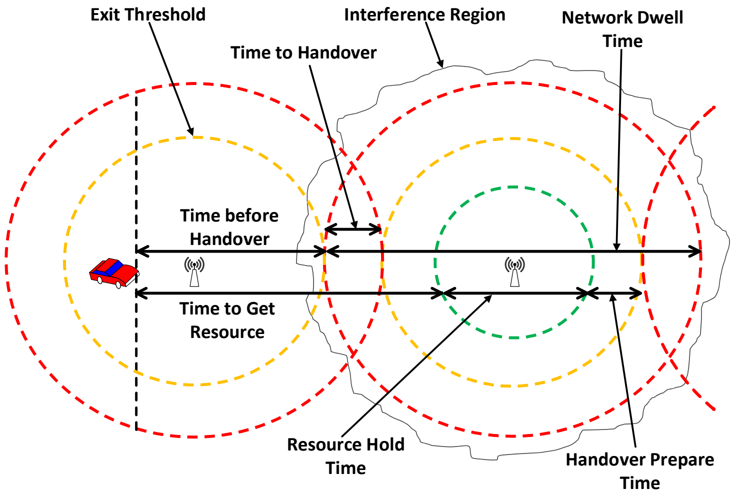


Figure 2: Communication range segmentation

Time ($\aleph$) as defined in Y-Comm Framework is the time MN will spend in the coverage i.e., the Network Dwell Distance (NDD) of new network. Time to get resource, $\mathbb{T}$ is the time when actual resource requested is available to the requested user i.e., even after entering the network's coverage range with a successful soft handover, the resource required by the MN might not be available, for example, other users might be holding the resource. Resource Hold Time, $\mathbb{N}$ is the resource usage time or when actual exchange of data is taking place. Handover Prepare Time (ϱ) is the time taken to prepare for handover during which the resource usage or data transmission will be paused and will be resumed after successful handover to the new network. Usually $\hbar$ and ϱ are very small compare to the values of other segments and therefore, $\mathbb{T}$ can be approximately equal to Υ if there are resources available in the new network, i.e., if there is no contention.

With the knowledge of these coverage parameters, it is possible to enhance the resource management based on the user contention for resources in a mobile network with proactive handover. A new proactive resource allocation based on the user contention to acquire a wireless channel resource is presented in the following section.

4 The proposed analytical model for proactive resource allocation

This section presents the proposed model to maintain QoS for using proactive resource management approach in highly mobile environments. In this paper, all the possibilities of contention interaction i.e., no contention, partial contention and full contention have been explored to show the practicability as well as the operational spaces of the overall system for highly mobile users in cellular system. These interactions may result in a request leaving the contention queue due to full contention with a subsequent request or the request in queue may be rearranged due to partial contention. The contention queue only queues requests before the MNs reach the next network. Once the MN reaches the relevant network, its channel request in the contention queue will be placed in the channel allocation queue. The proposed proactive resource allocation model is shown in Figure 3.

In the proposed approach, the decision algorithm decides whether the MN's request will be admitted to the system based on values of $\mathbb{T}$ and $\mathbb{N}$ of all the MNs requesting a channel. There are three possible contention possibilities that affect mobile users in this scenario. The probability of full contention before entering the queue is given as α. In addition, the probability of full contention occurring in the contention queue can be expressed as β. Finally, partial contention

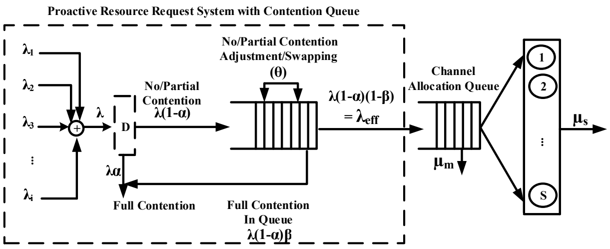


Figure 3: Proactive resource allocation queuing system with β and α

in the contention queue can be expressed as θ and this may result in modification of the request and/or re-ordering of request in the contention queue. Since θ does not involve in leaving the contention queue therefore, it will have no overall effect with respect to the rate of transfer to the channel allocation queue in this paper. The formulation of θ and validation is still in progress and will be considered as a future work.

4.1 The proposed proactive Markov queuing model

The proposed proactive resource allocation queuing model considers S number of channels and can allow i requests at time t. Q is the queueing capacity of the proposed system. The arriving requests may be sent from different MN to the system. It is assumed that the originating calls can join the system with an arrival rate of λ_O. Similarly, the handover calls can join the system with an arrival rate of λ_H. Hence, the total arrival rate is λ=λ_H+λ_O. According to [25], for a two-dimensional fluid model, the arrival rate of handover calls can be obtained as follows:

$$\lambda_H \approx \frac{\mu_m}{\mu_s}\lambda_O \tag{1}$$

Hence, the inter-arrival time of consecutive task follows the Poisson process which can be distributed as an exponential distribution with arrival rate λ=λ_H+λ_O.

$$\lambda = \sum_{i=1}^{S} \lambda_i \tag{2}$$

If there is no full contention in the contention queue, the mobile calls can join the system with an arrival rate of $\lambda(1-\alpha)$. Thus, scheduling and arrangement of requests take place and the total effective arrival rate can be calculated as:

$$\lambda_{eff} = \lambda(1-\alpha) * (1-\beta) \tag{3}$$

Figure 4 shows the state diagram of the proposed model. Let's define the states i (i=0,1,2,$\cdots$,S+Q) as the number of requests in the system at time t.

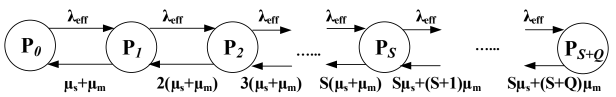


Figure 4: State diagram of the proposed approach

In proposed model, $\mathbb{T}$ and $\mathbb{N}$ are exponentially distributed with a mean rate of μ_s and μ_m, respectively. μ_m can be calculated based on the literature in [9, 10, 25]. Equation 4 is used in the literature for the dwell time in wireless and mobile systems [10,19,25] for handover queuing models. Thus, μ_m can be calculated and described as follows:

$$\mu_m = \frac{E[\nu] \cdot L}{\pi \cdot A} \tag{4}$$

where, $E[\nu]$ is the average velocity (ν) of MN, 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. Hence, the total channel holding time of a call is exponentially distributed with mean $1/(\mu_s + \mu_m)$. 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 $i(\mu_s + \mu_m)$ or $S\mu_s + i\mu_m$ if $S \leq i \leq S + Q$ as shown in Figure 4.

$$\mu_i = \begin{cases} i(\mu_s + \mu_m) & 0 \leq i < S \\ S\mu_s + i\mu_m & S \leq i \leq S + Q \end{cases} \tag{5}$$

ρ is the traffic intensity in the system, where $\rho = \lambda_{eff}/(\mu_s + \mu_m)$. As the requests are rejected from entering the system (α) and requests in the contention queue can also be removed (β) due to full contention with subsequent requests. Assuming a system in a steady state, the state probabilities, P_i's, can be obtained as in equation 6. P_i is the probability that there are i calls in the system.

$$P_i = \begin{cases} \frac{\rho^i}{i!} \cdot P_0 & 0 \leq i \leq S \\ \dfrac{\frac{\rho^S}{S!} \cdot \lambda_{eff}^{i-S} \cdot P_0}{\prod\limits_{j=S+1}^{i} [S\mu_s + (j-S)\mu_m]} & S < i \leq S + Q \end{cases} \tag{6}$$

In Equation 6, P_i is the probability that there are i calls in the system. P_0 can be defined as follows:

$$P_0 = \left[\sum_{i=0}^{S} \frac{\rho^i}{i!} + \sum_{i=S+1}^{S+Q} \frac{\frac{\rho^S}{S!} \cdot \lambda_{eff}^{i-S}}{\prod\limits_{j=S+1}^{i} [S\mu_s + (j-S)\,\mu_m]} \right]^{-1} \tag{7}$$

The mean queue length (MQL) i.e., the average number of requests in the system can then be calculated as $MQL = \sum\limits_{i=0}^{S+Q} i \cdot P_i$.

$$MQL = \left[\sum_{i=0}^{S} \frac{i\rho^i}{i!} + \sum_{i=S+1}^{S+Q} \frac{i \cdot \frac{\rho^S}{S!} \cdot \lambda_{eff}^{i-S}}{\prod\limits_{j=S+1}^{i} [S\mu_s + (j-S)\,\mu_m]} \right] P_0 \tag{8}$$

Similarly, the blocking probability (P_B), throughput (γ) and mean response time (MRT) of the system can be calculated as follows:

$$P_B = P(S+Q) = \frac{\frac{\rho^i}{S!} \cdot \lambda_{eff}^{Q} \cdot P_0}{\prod_{j=S+1}^{S+Q} \left[S\mu_s + (j-S)\,\mu_m\right]} \tag{9}$$

$$\gamma = \sum_{i=0}^{S+Q} i \cdot \mu_i P_i \tag{10}$$

$$MRT = \frac{MQL}{\gamma} \tag{11}$$

4.2 Request management in the contention queue

An example of the proactive resource allocation in queue is shown in Figure 5. Let us consider a simple queue and requests with ($\mathbb{T}$,$\mathbb{N}$) arrive to the queue as shown in Figure 5. Req_A arrives with (10,10) in seconds to the contention queue, the decision algorithm checks the queue and Req_A is queued at the front as the queue is empty.

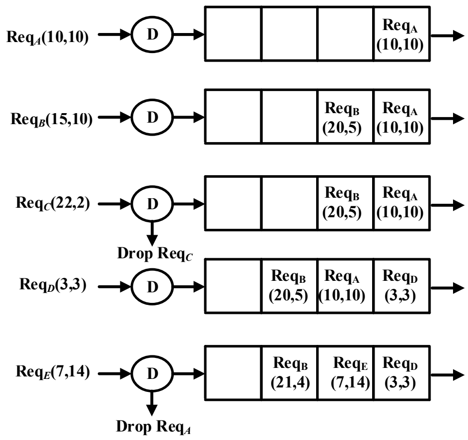


Figure 5: An example of proactive resource allocation management in queue

Now Req_B arrives with (15,10) and the time when it needs the channel and is when Req_A will still be using the channel resulting in partial contention for Req_B. Because, Req_A release the channel at the end of 20. Therefore, Req_B is modified to $\mathbb{T}_{\mathbb{B}} = 20$ and $\mathbb{N}_{\mathbb{B}}$ is modified to $(15+10)-20$ hence, the modified request for Req_B is (20,5). Req_C now arrives with (22,2) however Req_B will release the channel at 25 which means that Req_C will never get the channel and therefore it is not admitted to the contention queue due to full contention. A rejection reply is sent to Node C causing it to do an immediate handover to another network. Now Req_D arrives with (3,3), there is no contention with Req_A or Req_B and therefore, it is placed at the head of the queue since $\mathbb{T}_{\mathbb{C}} + \mathbb{N}_{\mathbb{C}} < \mathbb{T}_{\mathbb{A}}$ i.e., $6 < 10$. Req_E arrives with (7,14) this results in no contention with Req_D, full contention with Req_A and hence Req_A is ejected from the contention queue because $\mathbb{T}_{\mathbb{E}} < \mathbb{T}_{\mathbb{A}}$ and $\mathbb{T}_{\mathbb{E}} + \mathbb{N}_{\mathbb{E}} > \mathbb{T}_{\mathbb{A}} + \mathbb{N}_{\mathbb{A}}$. Req_B experiences partial contention and hence the request is modified accordingly.

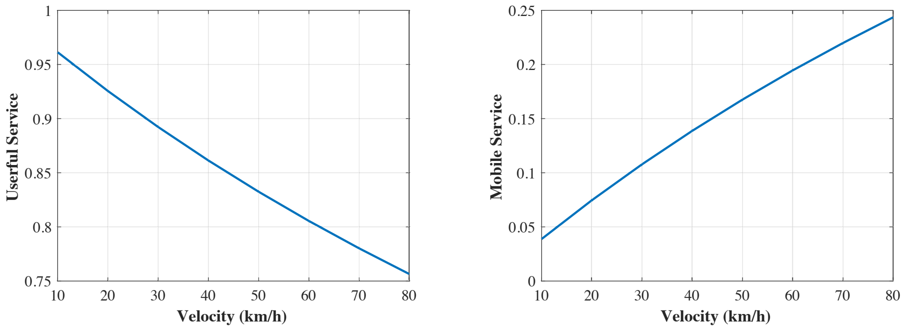
4.3 Useful service vs mobile service

Since the MN can leave the queuing system due to service (μ_s) or due to mobility (μ_m), it is necessary to distinguish these two events to properly reflect the performance of the system. We therefore define two concepts which are important in a mobile environment. The first is useful service in classical handover, U_{sc}, where the MN leaves the system after using the channel. When the MN leaves the system due to mobility and is not served by the channel, this is called Mobile Departure or Service, U_{mc}. For the classical case, we can represent these parameters as follows:

$$U_{sc} = \frac{S\mu_s}{S\mu_s + Q\mu_m}, \qquad U_{mc} = \frac{Q\mu_m}{S\mu_s + Q\mu_m} \tag{12}$$

We say that a system in which $U_{sc} > 0.5$ represents that 50% of the overall service rate is due to the channel being used, while $U_{sc} < 0.5$ represents under utilization because most of the overall service rate is due to MNs leaving the system due to mobility. These concepts are applied to give a better understanding of the effects of mobility.

The system parameters used are taken from [9] for consistency. The system has a fixed number of identical channels: $S = 12$. Q is the queuing capacity, which represents the number of requests waiting for service and is limited to 100. The service rate of the system μ_s is 0.01 requests/sec. The average speed of the MN and the radius of the network are taken as 10km/h to 80km/h and 1000m for all calculations, respectively. The rates are translated into requests per second in order to use consistent values.



(a) Useful service (b) Mobile service

Figure 6: Useful service vs mobile service

We can observe from Figure 6 that as the velocity of the MN increases, the useful service of the system is significantly reduced and the mobile service is increasing. This means that the number of MNs which are moving out of the network due to mobility without being served is increasing. Hence, in mobile networks it is necessary to take into account the mobility aspects to ensure effective resource management, so that we can optimize the useful service of the channels. In these environments most users will leave the system without being served resulting in poor user experience as well as poor system performance. The key observation is that because of the reactive approach in the classical model, MNs can queue for a channel with no hope of getting the channel while they are in the coverage area of the network. As a result this is useless waiting and if this inability to obtain a channel can be signalled to the MN before queueing for the channel then it would allow the user to immediately look for an alternative network and hence would improve the Quality of Experience (QoE) as well as the overall system performance. This means that the analysis of contention between different MNs has to be analysed in detail so that only MNs that have a chance of getting the channel should be queued for service.

5 Results and discussion

This section presents numerical results in order to show the accuracy and effectiveness of the proposed analytical model of the proactive approach to improve resource allocation in highly mobile environments. In addition, the results obtained are analysed and formulated for β and α by investigating all contention probabilities. The system parameters used are mainly taken from [9] and relevant literature [10, 19]. The system has a fixed number of identical channels: S=12 and the queue capacity is 100. 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 [9] where on average a minimum of 2 slots are 0.5 ms. Hence, the rates are translated into request per second in order to use consistent values. The service rate of the mobile users μ_m is calculated using Equation 4. The requests are entering the system due to the proposed contention system with arrival rate λ_{eff}.

5.1 Effects of α and β

The focus of the paper is to understand the effects of β and α by investigating all contention probabilities together with the operative area of various parameters to maintain QoS for resource management. In [9] all possible contention probabilities are considered with different scenarios. However, more accurate results of a request not being affected by β and α are analysed and presented. In addition, 3D graphs are also generated in order to see a complete picture for such systems. The parameters taken for the rest of the figures are; S=12, Q=100, μ_s=0,02(reqs/sec), $E[\nu]$=50km/s and R=1000m, unless stated otherwise.

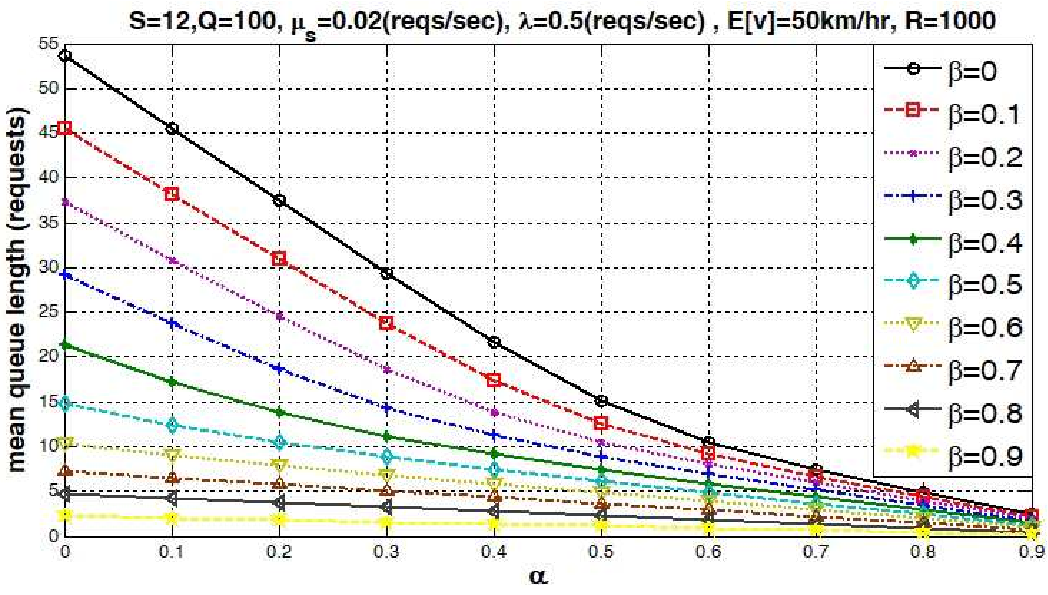
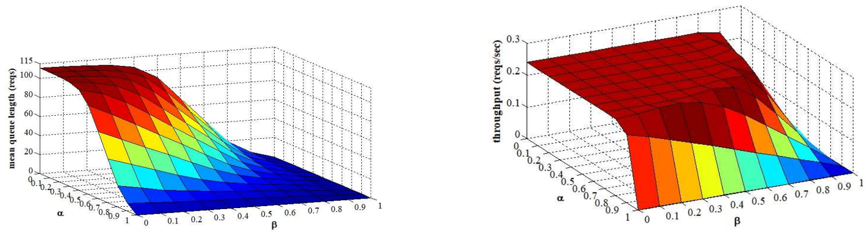


Figure 7: MQL results vs α with different β

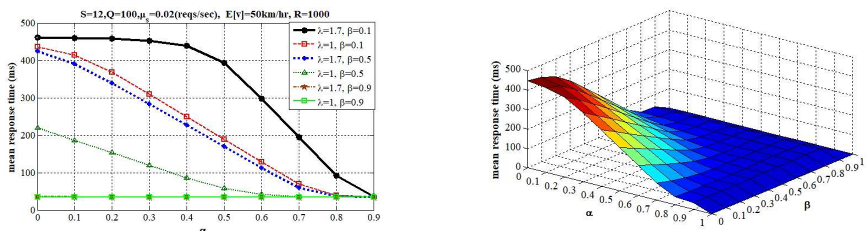
The figure 7 shows MQL results as a function of α considering all full contention probabilities occurring in the queue (β) for a light traffic loads (e.g, λ=0.5). The figure clearly shows the importance of β. Even though, the system has a light traffic loads, the MQL increases without considering β. However, the MQL decreases as β increases. This is due to the full contention occurring in the queue. MNs are not allowed to the system since they will not have enough time to get service. MNs carry on their journey without entering the system.



(a) Mean queue length (b) Throughput

Figure 8: 3D representation of all contention probabilities for λ=1.7

Hence, the operational area of the MNs in acquiring a resource is shown in Figure 8 for high traffic loads (e.g., λ=1.7) in 3D. For instance, the system is almost full (MQL=110.98 reqs/sec) when there is no contention (β=α=0). However, considering both contention probabilities the decision of acquiring a resource for MNs can be efficiently achieved by analysing the results in Figure 8. The Figure 8a shows that, the MNs start to line up when β=0.6 and α=0.7 for high traffic loads (e.g., λ=1.7). However, the MNs can get a channel even when the contention queue is fully loaded. In addition, when the full contention probability in the queue is high in the system, it affects the overall performance significantly. For instance, the MNs can get a channel when β=0.7 regardless of α. On the other hand, in order to obtain optimum QoS, the system performance depends also on α values for low values of β. The throughput results are shown in 3D in Figure 8b considering all possible contention probabilities. The results indicate the drop in the throughput of the given both traffic loads since more requests are removed from the system due to contention. In other words, the throughput decreases as MNs leave the system considering the contention probabilities. However, this improves the overall network performance as these requests can be dealt with by other networks.



(a) MRT vs α (b) 3D representation of MRT for λ=1.7

Figure 9: Mean response time results for all contention probabilities

The Figure 9a shows MRT results as a function of α considering various full contention probabilities occurring in the queue (β) for different traffic loads (e.g., λ=1, 1.7). As clearly seen from the figure, the best MRT results can be obtained for high value of β. When β=0.9, the all MNs can get a channel regardless of α. Even though the system has moderate traffic loads (e.g., λ=1) the full contention probability in the queue is important parameter of getting a resource in such system. This can be clearly seen in Figure 9b.

5.2 Probability of a request staying in the contention queue

In the queueing analysis presented in previous section, α and β values are assumed. In order to utilize the proposed proactive approach, a model has to be developed to obtain the probability of a request not been affected by α and β. It is known from the standard queueing analysis for a system of k nodes:

$$P_n = \rho^n \frac{1-\rho}{1-\rho^{k+1}} \tag{13}$$

Where, ρ is the traffic intensity (λ/μ), k is the size of queue and n is the number of request in the system.

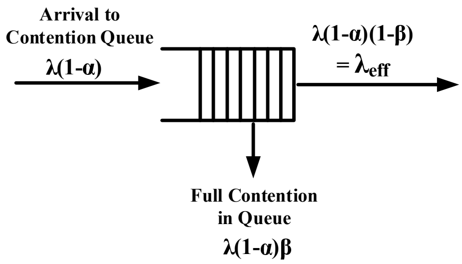


Figure 10: Proactive contention queue

It can be assumed that the proactive contention queue as a simple M/M/1/k queue. Here, we are interested only to find out the probability of a given request staying in the contention queue and not be affected due to β. Here the arrival rate to the contention queue is $\lambda(1-\alpha)$ and service rate is β as we are only interested in the requests affected by β as shown in Figure 10. Therefore, ρ can be represented as:

$$\rho = \frac{\lambda(1-\alpha)}{\beta} \tag{14}$$

Therefore,

$$P_{n-1} = \rho^{n-1} \frac{1-\rho}{1-\rho^{k+1}} \tag{15}$$

Here, the number of requests in the system becomes $n-1$ as there are no entries in the server. In order to find the probability of a given request staying in the contention queue without being affected by β, A request needs to satisfy three conditions:

- The new incoming request occupied the N^{th} position in the queue, i.e., last position of the queue.
- New incoming request is having a full contention with an entry at the N^{th} position and pushing that entry our of the queue.
- New request after occupying a place in the queue is not affected by the subsequent new incoming request.

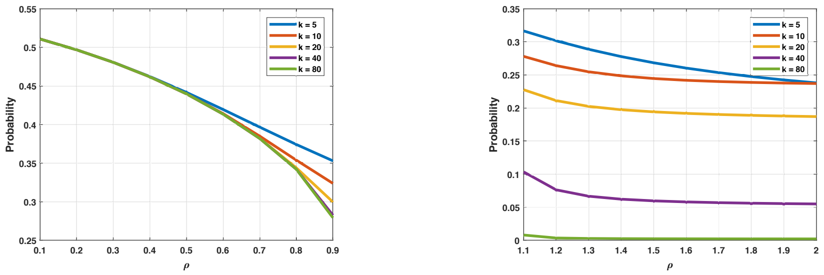
Here, let us assume that P^A_{Full} is the probability of full contention for incoming request A. P^B_{Full} is the probability of full contention for the requests in contention queue and here an average velocity is considered. P^C_{Full} is the probability of full contention for a new subsequent incoming request C after A.

- $(1-P^B_{Full})^{N-1}$ is the probability that request A occupies the N^{th} position in the contention queue and it is not colliding with any other request before it.
- $(1-P^A_{Full})^{N-1}P^A_{Full}$ is the probability that request A do not collide with the request in front of N^{th} position but does collide with the request already occupying the N^{th} position.
- $1-[(1-P^C_{Full})^{N-1}P^C_{Full}]$ is the probability of the request A not getting booted out by a new subsequent request entering the queue i,e C.

Therefore, all these three conditions can be represented as shown in Equation (16) to calculate the probability of a request staying in the queue and not been affected by β.

$$\sum_{n=1}^{n=k}\Big\{\Big[\Big(\frac{(\rho^{n-1})(1-\rho)}{1-\rho^{k+1}}(1-P^B_{Full})^{n-1}\Big)+\Big(\sum_{m=n}^{m=k}\frac{\rho^m(1-\rho)}{1-\rho^{k+1}}(1-P^A_{Full})^{m-1}P^A_{Full}\Big)\Big]\times \Big(1-(1-P^C_{Full})^{n-1}P^C_{Full}\Big)\Big\}\times(1-P^B_{Full}) \tag{16}$$

For example, let us consider $\mathbb{T}$ for MN_A is 10s and MN_B is 60s. $\mathbb{N}$ for MN_A is 20s and MN_B is 30s. The resulting probability of full contention based on our two node model: $P^A_{Full} = 0.085714$, $P^B_{Full} = 0.064286$ and let us assume $P^C_{Full} = 0.2$. Probability of a request not affected by β for $k = 1$ is 0.68635, $k = 20$ is 0.45405 and $N = 80$ is 0.75405. Calculating β accurately is very important so that the requests leaving contention queue due to full contention can be served by an alternative network. Since, β is dependent on the values of $\mathbb{N}$ and $\mathbb{T}$ of each request in the queue and new incoming request, therefore, a detailed analysis is required to accurately model β. By using this approach, it is possible to achieve seamless communication.



(a) $\rho = 0.1\ to\ 0.9$ (b) $\rho = 1.1\ to\ 2.0$

Figure 11: Probability of a request not being affected by β

Let us consider $\mathbb{T}_A$ and $\mathbb{T}_B$ are both 20s. MN_A is moving at a velocity of 30Mph and average velocity of the requests in the contention queue as 10Mph. Therefore, for network the P^A_{Full} is 0.094055 and P^B_{Full} which is the probability of full contention for the requests with average velocity in the queue is 0.346165. Let us assume P^C_{Full} as 0.2. Therefore, we can find the probability of request A staying in the queue using Equation (16) and the resulting graphs for different ρ values from 0.1 to 0.9 and ρ values from 1.1 to 2 are shown in Figure 11a and 11b, respectively. From the result we can observe that as ρ increases the probability of a request being affected by β, resulting in not reaching the channel queue increases. In other words, probability of a request not being affected by β deceases as shown in Figure 11. In addition, it is clearly seen from the Figures 11a and 11b that k affects the system performance significantly especially for highly loaded system.

6 Conclusion and future work

Analytical models for a proactive system where the users contend for resources with a thorough analysis has been presented. These models have yielded two key parameters, β and α, that have to be calculated to effectively use the proposed proactive approach. It has been shown that if these two parameters are identified then we can ensure that each users in the system will be effectively served by at least one of the available network. This can be achieved by using vertical handover to alternative network if an user is about to experience a full contention in the target network. To the best of our knowledge there is no study that has modelled a proactive model in the context of users contending for resources based on their mobility. With the results showing that the proposed approach is outperforming the classical model. The application of the work and approach in a Middlesex VANET testbed is the future work.

Bibliography

[1] Alhabo, M.; Li, Z.; Naveed, N. (2017); A Trade-off Between Unnecessary Handover and Handover Failure for Heterogeneous Networks. *European Wireless 2017*, 167-172, 2017.

[2] Almulla, M.; Wang, Y.; Boukerche, A.; Zhang, Z. (2014); Design of a fast location-based handoff scheme for IEEE 802.11 vehicular networs, *IEEE Transactions on Vehicular Technology*, 63(8), 3853-3866, 2014.

[3] Bejaoui, T. (2014); QoS-oriented high dynamic resource allocation in vehicular communication networks, *The Scientific World Journal*, Article ID 718698, 9 pages, 2014. http://dx.doi.org/10.1155/2014/718698

[4] Chen, X.; Ni, M. (2016); Seamless handover for high mobility environments, *International Wireless Communications and Mobile Computing Conference (IWCMC)*, 281-286, 2016.

[5] Elhadj, H. B.; Elias, J.; Chaari, L.; Kamoun, L. (2016); Multi-attribute decision making handover algorithm for wireless body area networks. *Comput. Commun.*, 81, 97-108, 2016.

[6] Feirer, S.; Sauter, T. (2017); Seamless handover in industrial WLAN using IEEE 802.11k, *IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 1234-1239, 2017.

[7] Fernandez, P. J.; Santa, J.; PerenÄąguez, F.; Skarmeta, A. F. (2017); Towards seamless inter-technology handovers in vehicular fIPv6g communications, *Computer Standards and Interfaces*, 52, 85-96, 2017.

[8] Huang, Q.; Huang, Y.-C.; Ko, K.T.; Iversen, V.B. (2011); Loss performance modeling for hierarchical heterogeneous wireless networks with speed-sensitive call admission control, *IEEE Transactions on Vehicular Technology*, 60(5), 2209-2223, 2011.

[9] Kirsal, Y.; Paranthaman V.V.; Mapp,G. (2018); Exploring analytical models to maintain quality-of-service for resource management using a proactive approach in highly mobile environments, *2018 7th International Conference on Computers Communications and Control (ICCCC), Proc. of,* Oradea, Romania, May 2018, Publisher: IEEE, 176-182, 2018. DOI: 10.1109/ICCCC.2018.8390456

[10] Kirsal, Y. (2016); Analytical modelling of a new handover algorithm for improve allocation of resources in highly mobile environments, *International Journal of Computers Communications & Control*, 11(6), 789-803, 2016.

[11] Li, J.; Luo, T.; Ding, L. (2014); A location aware based handoff algorithm in v2i system of railway environment. *Global Information Infrastructure and Networking Symposium (GIIS)*, 1-6, 2014.

[12] Mapp, G.; Katsriku, F.; Aiash, M.; Chinnam, N.; Lopes, R.; Moreira, E.; Vanni R.M.P.; Augusto, M. (2012); Exploiting location and contextual information to develop a comprehensive framework for proactive handover in heterogeneous environments. *Journal of Computer Networks and Communications*, 1-17, 2012.

[13] Mapp, G. E.; Cottingham, D.; Shaikh, F.; Vidales, P.; Patanapongpibul, L., Baliosian, J.; Crowcroft, J. (2006); An architectural framework for heterogeneous networking, *International Conference on Wireless Information Networks and Systems*, 2006.

[14] Nadembega, A.; Hafid, A.; Taleb, T. (2015); A destination and mobility path prediction scheme for mobile networks, *IEEE transactions on vehicular technology*, 64(6), 2577-2590, 2015.

[15] Paranthaman, V.V.; Ghosh, A.; Mapp, G.; Iniovosa, V.; Shah, P.; Nguyen, H. X.; Gemikonakli, O.; Rahman, S. (2017); Building a Prototype VANET Testbed to Explore Communication Dynamics in Highly Mobile Environments, *Springer International Publishing*, 81-90, 2017.

[16] Paranthaman, V. V.; Kirsal, Y., Mapp, G.; Shah, P.; Nguyen, H. X. (2017); Exploring a New Proactive Algorithm for Resource Management and Its Application to Wireless Mobile Environments. *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, 539-542, 2017.

[17] Saxena, N.; Roy, A. (2009); Experimental framework of proactive handover with qos over wlans. *Electronics Letters*, 45(25), 1313-1315, 2009.

[18] Shaikh, F.;, Mapp, G. E.; Lasebae, A. (2007); Proactive policy management using tbvh mechanism in heterogeneous networks, *International Conference on Next Generation Mobile Applications, Services and Technologies*, 151-157, 2007.

[19] Trivedi, K. S.; Dharmaraja, S.; Ma, X. (2002); Analytic modeling of handoffs in wireless cellular networks, *Information sciences*, 148(1), 155-166, 2002.

[20] Vidales, P.; Baliosian, J.; Serrat, J.; Mapp, G.; Stajano, F.; Hopper, A. (2005); Autonomic system for mobility support in 4g networks, *IEEE Journal on Selected Areas in Communications*, 23 (12), 2288-2304, 2005.

[21] Vilaplana, J., Solsona, F.; Teixido, I.; Mateo, J.; Abella, F.; Rius, J. (2014); A queuing theory model for cloud computing. *J. Supercomput.*, 69(1), 492-507, 2014.

[22] Wang, W.; Xu, H.; Zhou, H. (2014); Proactive unnecessary handover avoidance scheme in LTE-A small cells, *IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, 2214-2218, 2014.

[23] Wu, S.J. (2010); An intelligent handover decision mechanism for heterogeneous wireless networks, *The 6th International Conference on Networked Computing and Advanced Information Management*, Seoul, 688-693, 2010.

[24] Xenakis, D.; Merakos, L.; Passas, N.; Verikoukis, C. (2016); Handover Decision for Small Cells: Algorithms, Lessons Learned and Simulation Study. *Communication Networks*, Elsevier, 100, 64-74, 2016.

[25] 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, 2001.