An Analytical Approach for Performance Analysis of Handoffs in the Next Generation Integrated Cellular Networks and WLANs

Y. Kirsal, E. Gemikonakli
Computer Communication Department
Middlesex University
London, UK
y.kirsal, e.gemikonakli@mdx.ac.uk

E. Ever, G. Mapp, and O. Gemikonakli
Computer Communication Department
Middlesex University
London, UK
e.ever, g.mapp, o.gemikonakli@mdx.ac.uk

Abstract — The main feature of the next generation wireless communication systems is the ability to establish ubiquitous and seamless access to various radio access technologies (RATs) and standards. For this reason the integration of cellular and wireless local area networks (WLANs) and performance evaluation of the interaction between these technologies is now an important research area. Modelling such systems for performance evaluation is essential to improve the architecture according to the quality of service (QoS) requirements and performance characteristics. In this paper, an analytical model for performance evaluation of an integrated cellular network and a WLAN is considered. WLAN is deployed inside of the cellular network to support handoffs between cellular networks with higher bandwidth. Such an integrated system can be modelled as a two stage open network. An analytical model is proposed together with an exact solution technique in order to evaluate the performance of an integrated system consisting of a cellular network and a WLAN. A two stage queuing system is considered for this purpose. Numerical results are presented for mean queue length values of cellular system as well as the WLAN.

Keywords-component; integrated cellular/WLAN; performance analysis; handoff schemes; spectral expansion

I. INTRODUCTION

The next generation wireless systems, referred to as heterogeneous wireless networks, are expected to support different radio access technologies, high-rate multimedia services and ubiquitous access anytime and anywhere. Dealing with quality of service requirements, and providing the best configuration in order to maximise the performance characteristics of such systems is a challenging task mainly because of the heterogeneous nature of next generation wireless communication systems. In order to achieve these goals, integrating various technologies is an effective way. Convergence technology is a mechanism that attempts to seamlessly integrate various existing wireless networks instead of developing a new uniform standard for wireless communication systems [4]. Eventually, such networks will provide integrated and seamless services.

Integration of cellular network technologies with WLANs is a common configuration especially when third generation cellular networks are considered. However, in this paper we employ a generic approach in order to analyse the interaction between WLANs and cellular networks. Cellular networks have been well deployed around the world. They support both circuit-switched and packet-switched services. A well-defined infrastructure is the main benefit in cellular networks and has a global coverage area. However, cellular networks usually have low data rates and capability to serve applications with potential high bandwidth requirements such as video conferencing, and high-resolution image applications. On the other hand, WLANs usually support packet-switched services, and moreover these technologies have smaller coverage area while providing higher data rates at lower costs. WLAN technologies are efficient in serving higher data traffic but they have limitations such as poor mobility management [2], [4], [15].

Most of the previous research studies consider homogeneous cellular networks without considering the characteristics of WLANs [5], [6], [9], [10], [13], [14] and [17]. In other words, the integrated cellular/WLAN structure is not considered with an exact solution technique for performance evaluation. WLANs are generally deployed in indoor environments such as homes, offices, cafes, etc. and they are now seen as a complementary technology that can be used to enhance the system capacity in integrated wireless networks. In order to support integrated cellular/WLAN systems, provide an acceptable level of QoS and maximize the system capacity, modelling and performance analysis of such systems has to be taken into account. In this study, the modelling and performance analysis focus on the deployment of WLAN access point which is located inside a cellular network.

One of the main issues in determining the network performance in integrated networks is mobility. In [5], [6] and [10] horizontal handoffs, and mobility between homogeneous networks are considered (e.g. cellular to cellular). However, users experience vertical as well as horizontal handoffs within an integrated (heterogeneous) system. Because of the heterogeneous nature of modern communication systems, the network engineers should consider the best configuration where vertical handoffs are considered together with horizontal ones.

Various interworking and handoff strategies for the heterogeneous networks have been presented in [1], [2], [4], [15] and [16]. The main attention has been given to the interworking architecture, channel assignment schemes and/or mobility. Mobility issues such as the speed of mobile users are not taken
into account for both cellular and wireless local area networks, together with various radius characteristics. In this paper, an analytical model is developed for performance evaluation of integrated cellular/WLAN systems. Mobility issues are considered for both cellular and wireless local area network and an exact solution method is employed in order to analyse the interactions between two integrated systems and to provide a tool for analysing performance characteristics. Also the proposed model takes horizontal and vertical handoffs, user mobility, queue capacity and characteristics of cellular network and WLAN into account. The proposed model shows that such an integrated system can be modelled as a two stage queuing system. Using this model, important performance metrics, such as mean queue length can be computed. The spectral expansion method is employed for the exact solution of the analytical model considered [8].

Various channel assignment strategies such as fixed channel assignment (FCA), dynamic channel assignment (DCA) are available [3], [13], and FCA is assumed in this paper where the set of channels are permanently assigned to both base station and access point for cellular network and WLAN respectively [5], [6].

The remaining sections of the paper are organised as follows: Section II describes the proposed model and states the assumptions used. The two dimensional modelling approach and exact steady state solution are explained in section III. In section IV, the numerical results are presented which are computed by using an exact solution approach. Conclusions and recommendations are provided in section V.

II. THE SYSTEM AND THE ANALYTICAL MODEL

The proposed model considers the modelling of the heterogeneous wireless systems for performance evaluation. As shown in Fig. 1 the WLAN is deployed inside a single cellular network. The cellular network has a larger coverage area than the WLAN. Because the WLAN is within the coverage area of the cellular network, this provides the opportunity for users on the cellular to switch over to the WLAN or remain connected to the cellular network. It is assumed that, in a particular time, each user can send one request to only one network for connection [1], [2], [4], [7], [8], [15] and [16]. If the channels are not available in cellular network and/or WLAN systems, the calls start to queue up.

All calls arrive in cellular network and WLAN have inter arrival times which follows Poisson distribution similar to the studies presented in [1], [2], [4], [7], [8], [15] and [16]. If the channels are not available in cellular network and/or WLAN systems, the calls start to queue up.

The ongoing calls in the cellular network can be moved towards the WLAN coverage area or neighbouring cells while they are either in the queue (due to mobility) or being served in the system. If a user moves to the WLAN coverage area, it is likely that a vertical handoff should proceed. If the WLAN queue is full, the new job request will be blocked.

Let \( \lambda_T \) be defined as the total arrival rate of calls in the cell. Then,

\[ \lambda_T = \lambda_c + \mu_{vh,w} \text{ where}, \lambda_c = \lambda_{oc,c} + \lambda_{hh,c} \]

(1)

On the other hand, for ongoing calls moving out of the WLAN coverage area a vertical handoff should proceed. Users in cell with low mobility are admitted to WLAN and the downward vertical handoff will proceed \( (\mu_{vh,c}) \). In this study the probability of vertical handoffs \( \text{(cell to WLAN or WLAN to cell)} \) are taken as standard uniformly distributed probabilities and used as input to the queuing system in order to investigate the effects of vertical handoffs on overall system performance. The WLAN users try to find a WLAN channel to use. They may move to the cellular system due to mobility. Blocking takes place only if cellular and/or WLAN cannot offer any waiting room for the incoming requests (when the queues are full).

The integrated cellular network and WLAN can be thought as a two stage open queuing system as shown in Fig. 2. In [7] and [8] similar modelling approaches are introduced for performance and performability evaluation of two stage open queuing systems. However open queuing systems with multiple servers in both stages are not considered. In the cellular coverage area, new originating calls with arrival rate \( \lambda_{oc,c} \), horizontal handoff calls between neighbouring cells with \( \lambda_{hh,c} \) and upward vertical handoff calls \( \mu_{vh,w} \) from WLAN coverage area can request admission from the cell. Similarly, new originating calls with rate \( \lambda_{oc,w} \) and downward vertical handoff calls with \( \mu_{vh,c} \) from cell can request admission from the WLAN.
in cellular network ($K > S$) [1], [2]. For both networks, the maximum number of calls allowed into the system is equal to the number of calls assigned with the channels. The queuing capacity of the cellular system is given by $L_c$ where $L_c = S + W_c$. Similarly, in WLAN the capacity is given by $L_w$ where $L_w = K + W_w$. For analytical tractability, the hexagon cell shape and WLAN have circular shape of radiuses of $R$ and $r$, respectively. $T_{oc,c}$ and $T_{oc,w}$ are the average call holding time in cellular and WLAN which follows an exponential distribution with means $E[T_{oc,c}] = 1/\mu_{oc,c}$ and $E[T_{oc,w}] = 1/\mu_{oc,w}$ respectively. The cell dwell time, which is the time mobile stations (MSs) spend in the cell before moving to neighbouring cells and passing through to in WLAN are assumed to be exponentially distributed with means $E[T_{cd,c}] = 1/\mu_{cd,c}$, $E[T_{cd,w}] = 1/\mu_{cd,w}$, respectively [5], [10]. The residence time in WLAN is also assumed to be exponentially distributed with mean $E[T_{wd}] = 1/\mu_{wd}$ [5], [10]. $\mu_{cd,c}$ and $\mu_{cd,w}$ are the departure rate of the calls from the cell to the neighbouring cells and cell to the WLAN respectively. Similarly, the departure rate of the residential calls from WLAN to cell (WLAN users) is $\mu_{owd}$, $\mu_{owd,c}$, $\mu_{owd,w}$ and $\mu_{wd}$ assumed to be exponentially distributed [5], [10] and they can be calculated by using the approach presented in [5] as follows:

$$\mu_{cd,c} = \frac{E_v[v]P_v}{\pi A_c} , \mu_{cd,w} = \frac{E_v[v]P_v}{\pi (A_c - A_w)} , \mu_{wd} = \frac{E_v[v]P_v}{\pi A_w} (2)$$

$E[v]$ is the average of the random variable, $V$ which is the speed of the mobile users, $P_v$, $P_w$ are the length of the perimeter of cell and WLAN and $A_c, A_w$ are the area of the cell and WLAN respectively. As shown in Fig. 2, if the calls occupy the channels, they may be transferred to the WLAN with a probability of $\theta_1$, be completed or they may leave the cell with a probability of $(1-\theta_1)$. In other words $\theta_1$ and $(1-\theta_1)$ are the probability of a call to be transferred from the cell to the WLAN and probability of a call to leave the cell, respectively. On the other hand, if the calls are in the queue and waiting a channel to be served, they may be transferred to the WLAN due to the mobility with probability of $\theta_2$, $(1-\theta_2)$ is assumed to be the probability of a call to stay in the cell until it gets a channel to be served. Let $\mu_c$ be the total service rate of completed call departures from a cell where,

$$\mu_c = \mu_{oc,c}(1-\theta_1) + \mu_{cd,c} (3)$$

Then, $\mu_T,c$ can be defined as the total service rate of cellular system where,

$$\mu_{T,c} = (\mu_{oc,c} + \mu_{vh,c} + \mu_{vh,w} + \mu_{oc,w} + \mu_{cd,c} + \mu_{cd,w}) \theta_1 \theta_2 (4)$$

On the other hand, the calls served in the WLAN with a mean rate of $\mu_{oc,w}$ leave the system with a probability of $(1-\theta_2)$, $(\mu_w = \mu_{oc,w}(1-\theta_2))$ or upward vertical handoff from WLAN to cellular network takes place with a probability of $\theta_2$. Then $\mu_{T,w}$ can be defined as the total service rate of WLAN system where,

$$\mu_{T,w} = \mu_w + \mu_{vh,w} + \mu_{vh,c} \theta_2 \theta_2 + \mu_{wd} (5)$$

### III. Two Dimensional Modelling and Steady State Solution of The Proposed Model

The state of the system at time $t$ can be described by a pair of integer valued random variables, $I(t)$ and $J(t)$, specifying the number of calls present at a time $t$ for the cell and the WLAN, respectively.

#### A. The Two Dimensional Representation

The model assumes that $I(t) = 0, 1, ..., L_c$, $t \geq 0$, is an irreducible Markov process, which represents the number of calls in the cellular network including the one(s) in service. $J(t) = 0, 1, ..., L_w$ is the total number of calls in the WLAN system at time $t$, including the one(s) in service. Then, $Z = \{(I(t), J(t)): t \geq 0\}$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, \{0, 1, ..., $L_c$\} x \{0,1, ..., $L_w$\}. The irreducible Markov process is used for performance evaluation of the two stage open network considered. Unlike similar studies presented [7], [8], [12] multiple servers are considered for both stages.

![Figure 3. The state diagram of the irreducible Markov process considered.](image-url)
assigned to incoming calls and the calls in the queue have the number of available channels, all of the available channels are On the other hand if the number of calls is greater than the minimum of number of calls and number of available channels. call. Therefore the downward transition rate is chosen as the number of available channels, a channel is assigned for each call. Therefore the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as $\mu_{oc}$ [10]. The matrix $C$ is defined below:

$$C = \begin{bmatrix}
K\mu_w & (K\mu_w, \theta_2 + (L_w - K)\mu_{oc}) & 0 & 0 \\
0 & K\mu_w & \ddots & 0 \\
0 & 0 & \ddots & (K\mu_{oc}, \theta_2 + (L_w - K)\mu_{oc}) \\
0 & 0 & 0 & K\mu_w \\
\end{bmatrix}$$

The elements of matrices $C$ and $C_j$ depend on the parameters $K$, $\mu_w$, $\mu_{oc}$, $\mu_{ocw}$, $L_w$, and $\theta_2$. The matrix $C$ is dependent on the number of calls for $j=0, 1, \ldots, L_w$. Therefore, the threshold $M$ is taken as $M=L_w$. If the number of calls in the system is less than the number of available channels, a channel is assigned for each call. Therefore the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as $\mu_{oc}$ [10]. The matrix $C$ is defined below:

$$C = \begin{bmatrix}
\mu_{ocw} & j\mu_{ocw}, \theta_2 & 0 & 0 \\
0 & \mu_{ocw} & \ddots & 0 \\
0 & 0 & \ddots & j\mu_{ocw}, \theta_2 \\
0 & 0 & 0 & \mu_{ocw} \\
\end{bmatrix}$$

for $0 \leq j \leq K$

$$C_j = \begin{bmatrix}
\mu_{ocw} & (\mu_{ocw}, \theta_2 + (j-K)\mu_{oc}) & 0 & 0 \\
0 & \mu_{ocw} & \ddots & 0 \\
0 & 0 & \ddots & (\mu_{ocw}, \theta_2 + (j-K)\mu_{oc}) \\
0 & 0 & 0 & \mu_{ocw} \\
\end{bmatrix}$$

for $K < j \leq K+W_w$

This system can be solved and the steady state probabilities, $p_{i,j}$, can be obtained using the steady state solution presented in the next section.

### B. Steady State Solution

The solution is given for a bounded queue (i.e. $S \leq L_c$ and $K \leq L_w$). Spectral expansion solution is employed and the details of the method used can be found in [7], [8], and [12]. However, the balance equations defined for $(M=1) \leq j \leq (L_w \times 2)$ are not used since the threshold value is defined as $M=L_w$. From the $p_{i,j}$ a number of steady-state performance measures can be computed quite easily. For illustration, we have concentrated on the mean queue length (MQL) of cellular system and WLAN which can be obtained as:

$$MQL = \sum_{j=0}^{L} \sum_{i=0}^{L} p_{i,j}$$

### IV. NUMERICAL RESULTS AND DISCUSSIONS

Performance measures have been computed using the model and spectral expansion solution approach. The values of mean arrival and mean service rates are mainly application dependent. In Fig 4. and Fig 5. cellular network with $S=10$ and $W_w=10$ are considered ($L_c=S+W_c=20$). The other parameters are $E[T_{oc,c}]=E[T_{oc,w}]=60$sec, $\lambda_{oc,c}$=0.01calls/sec, $E_c[v]$=40km/h [9], $E_w[v]$=2km/h, $R$=1000m, $r=100m$ [1],[2], $\theta_1=10\%$, $\theta_2=1%$ and $\theta_3=10\%$. Performance measure, mean queue length (mql), has been computed for both cellular system and WLAN ($MQL_c$ and $MQL_w$ respectively). The thinner line is used to represent $MQL_c$ where the thicker one is used for $MQL_w$.

![Figure 4. The MQL results for WLAN and cellular network with various K and W_w values.](image-url)
than the capacity of the cellular system. The mobile nodes accumulated in the queue of the WLAN leave the system depending on the radius of WLAN and the velocity of the nodes. Apart from the handoffs coming from the servers of the WLAN the handoffs due to mobility are fed into the cellular system. In this case since the capacity of the cellular system is much smaller (S=10, Wc=10 compared to K=50, Ww=50) the cellular system can become full quite rapidly, and makes the overall cellular/WLAN integration unstable. It is essential to provide a better configuration, or develop new algorithms for similar cases since the system is stable when $\theta_1$ value is around 1%. This means that most of the calls switching to the WLAN actually need to be completed in the WLAN without the need to the cellular network. This is more likely with data traffic which can use the aggregated bandwidth of the WLAN to complete its transfer.

On the other hand, the results presented in Fig. 5 illustrate that the cellular system is stable and $\theta_2$ value can have values up to 20% when $K=20$, and $W_w=20$. Please also note that the mean queue length of the WLAN is not affected significantly by the changes in the $\theta_2$ value. The calls in the queue of the cellular network may be transferred to the WLAN and the transition rates are increased dramatically. This behaviour can also affect the stability of the entire model. Thus, the $\theta_2$ value is another important parameter for design and optimisation of integrated systems considered.

V. CONCLUSIONS

The analysis of the handoffs is an important issue in order to get good performance from the next generation integrated cellular/WLAN networks. In this paper a generic analytical model in cellular/WLAN integration for heterogeneous wireless networks is considered. Unlike the previous studies, for such systems, an exact spectral expansion solution approach is used for the performance evaluation. The mobility issues such as, mobile users velocity in cellular and WLAN are considered separately.

Obviously the analysis should be carried even further for more informed decisions, since the interaction between WLAN and cellular system is essential for optimization studies of heterogeneous wireless communication systems. Furthermore potential channel failures should also be considered. Current solution approaches suffer from the well known state space explosion problem when such complicated scenarios are considered. The solution method provided in [7] looks promising, however because of the multiple servers used at both of the stages of the open network considered, it is desirable to develop alternative approaches to cope with the state space explosion problem and consider performability measures.

REFERENCES


