User-Network Association in an 802.11 WLAN & 3G UMTS Hybrid Cell: Individual Optimality

Dinesh Kumar, Eitan Altman
INRIA, France
Email: {dkumar, altman}@sophia.inria.fr

Jean-Marc Kelif
France Telecom R&D, France
Email: jeanmarc.kelif@orange-ft.com

***Abstract*— More and more users subscribing to wireless broadband services are seeking to have access to both WLAN and UMTS networks. We study individually optimal user-network association in an integrated WLAN and UMTS *hybrid cell*. The association problem is formulated within a non-cooperative game framework. In the formulation, mobile arrivals are assumed to follow the Poisson process and each mobile considers its average service time in each network as the decision criteria to connect to either of the WLAN or UMTS networks. We seek to compute the optimal association or decision policy that achieves the Nash equilibrium. For this we develop a generic system of linear equations for estimating the average service time of a mobile. This system is then solved assuming a particular model for the WLAN and UMTS networks and we explicitly compute the optimal association policy that is observed to possess a descending staircase curve structure.**

I. Introduction

As 802.11 WLANs and 3G UMTS cellular coverage networks are being widely deployed, network operators are seeking to offer seamless and ubiquitous connectivity for wireless broadband services through integrated WLAN and UMTS hybrid networks. One of the core decision problems faced in such a hybrid network is that of user-network association, i.e., decision of an arriving mobile user to connect to one of the two constituent networks. We study this decision problem in the framework of individual optimality where arriving mobile users selfishly connect to one of the two networks (WLAN or UMTS) based on an individual decision cost criteria. This gives a non-cooperative game structure to the decision problem and we compute the Nash equilibrium achieving optimal policy assuming a specific network model for the WLAN and UMTS networks. To be more precise consider a hybrid network, comprising two independent 802.11 WLAN and 3G UMTS networks, that offers connectivity to mobile users arriving in the combined coverage area of these two networks. By independent we mean that transmission activity in one network does not create interference in the other. Our goal in this paper is to study the dynamics of individually optimal user-network association in such a WLAN-UMTS hybrid network. An alternate approach based on globally optimal user-network association is envisaged to be studied as part of our future work. Note that we do not propose a full fledged cell-load or interference based connection admission control (CAC) policy in this paper. We instead assume that a CAC precedes the association decision control. A connection admission decision is taken by the CAC controller before any mobile is allowed to connect to either of the WLAN or UMTS networks. Thereafter, the mobile's association decision only optimizes its individual performance and it is not proposed as an alternative to the CAC decision.

A. Related Work and Contributions

Study of WLAN-UMTS hybrid networks is an emerging area of research and not much related work is available. Authors in some related papers [1], [2], [3], [4], [5], [6], [7] have studied issues such as vertical handover and coupling schemes, integrated architecture layout, radio resource management (RRM) and mobility management. However, questions related to user-network association have not been explored much. Premkumar et al. in [8] propose a *near optimal* solution for a hybrid network within a combinatorial optimization framework, which is different from our approach. To the best of our knowledge, ours is the first attempt to present a generic formulation of the user-network association problem under a non-cooperative game framework. Moreover, this work is the first we know of that obtains an explicit threshold based policy for the WLAN-UMTS hybrid network model that we consider.

II. Framework for the Decision Control Problem

A hybrid network may be composed of several 802.11 WLAN Access Points (APs) and 3G UMTS Base Stations (NodeBs) that are operated by a single network operator. However, our focus is only on a single pair of an AP and a NodeB that are located sufficiently close to each other so that mobile users arriving in the combined coverage area of this AP-NodeB pair have a choice to connect to either of the two networks. We call the combined coverage area of a single AP cell and a single NodeB micro-cell [12] as a *hybrid cell*. The cell coverage radius of a UMTS micro-cell is usually around $400m$ to $1000m$ whereas that of a WLAN cell varies from a few tens to a few hundreds of meters. Therefore, some mobiles arriving in the hybrid cell may only be able to connect to the NodeB, either because they fall outside the transmission range of the AP or they are equipped with only 3G technology electronics. While other mobiles that are equipped with only 802.11 technology can connect exclusively to the WLAN AP. Apart from these two categories, mobiles equipped with both 802.11 WLAN and 3G UMTS technologies can connect to any one of the two networks.

The decision to connect to either of the two networks may involve a cost criteria based on the average service time of a mobile in the hybrid network. Moreover, the connection or association decision involves two different possible decision

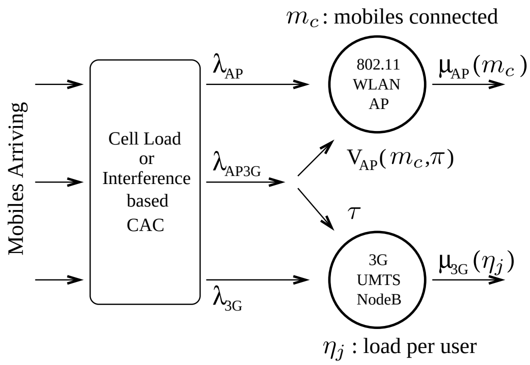

Fig. 1. Hybrid cell scenario

makers, the mobile user and the network operator. We focus only on the individually optimal control problem in which the mobile users take a selfish decision to connect to one of the two networks while optimizing only their own individual costs. In Section III, we motivate a non-cooperative game formulation for this problem. Our game formulation is a *generic* formulation of the user-network association problem in a WLAN-UMTS hybrid network and is independent of the network model assumed for WLAN and UMTS networks. Thereafter in Section V, we solve the problem assuming a particular network model described in Section IV.

A. Mobile Arrivals

We model the hybrid cell of an 802.11 WLAN AP and a 3G UMTS NodeB as a two-server processing system (Figure 1) with each server having a separate finite capacity of M_{AP} and M_{3G} mobiles, respectively. For simplification we assume that mobile users are stationary, having no mobility. As discussed previously, mobiles are considered as candidates to connect to the hybrid cell only after being admitted by a CAC such as the one described in [9]. Assuming mobile arrivals to be Poisson it can be easily shown using the PASTA property (Poisson arrivals see time averages) that mobile arrivals after the CAC are also Poisson with a reduced effective arrival rate. The reduced effective arrival rate can be obtained by incorporating the fraction of mobiles that are dropped by the CAC. Some of the mobiles (after they have been admitted by the CAC) can connect only to the WLAN AP and some others only to the UMTS NodeB. These two set of mobiles (or sessions) are each assumed to constitute two separate dedicated arrival streams with effective Poisson rates λ_{AP} and λ_{3G}, respectively. The remaining set of mobiles which can connect to both networks form a common arrival stream with effective Poisson rate λ_{AP3G}. The mobiles of the two dedicated streams directly join their respective AP or NodeB network without any connection decision choice involved. Mobiles of the common stream decide to connect to one of the two networks while optimizing their own cost.

B. Service Requirements and Departure Rates

It is assumed that all arriving mobiles have a downlink data service requirement which is exponentially distributed with parameter ζ. In other words, every arriving mobile seeks to download a data file of average size $1/\zeta$ bits on the downlink. Let $\theta_{AP}(m_c)$ denote the downlink packet (or file) throughput of each mobile in the AP network when m_c mobiles are connected to it at any given instant. If η_j denotes the *load factor* of a mobile j in the NodeB cell (see Chapter 8 in [12]) then $\theta_{3G}(\eta_j)$ denotes the downlink packet (or file) throughput of this mobile in the NodeB network. With these notations, the effective departure rates of mobiles (or sessions) in each server (or network) can be denoted by,

$$\mu_{AP}(m_c) = \zeta \times \theta_{AP}(m_c) \tag{1}$$

and

$$\mu_{3G}(\eta_j) = \zeta \times \theta_{3G}(\eta_j), \tag{2}$$

where, $\mu_{AP}(m_c)$ is identical for each mobile in the AP network and $\mu_{3G}(\eta_j)$ is different for each mobile j in the NodeB network and is a function of its load factor, η_j. The load factor, η_j, in turn depends on the location of mobile j in the NodeB cell.

III. Non-cooperative Game Formulation

As discussed earlier, a mobile arriving in the common stream selfishly decides to join one of the two networks so that its own cost is optimized. We consider the *average service time* of a mobile as the decision cost criteria and an incoming mobile connects to either the AP or NodeB network depending on which of them offers the minimum average service time. Service time here represents the time required for a mobile to accomplish its file download. Therefore, higher is the packet throughput achieved by a mobile, lesser will be its service time. We develop this model as an extension to the framework of [17] where an incoming user can either join a shared server with a PS service mechanism or any of several dedicated servers. Based on the estimate of its expected service time on each of the two servers, a user takes a decision to join the server on which its expected service time is least. This framework can be readily applied to our hybrid cell scenario so that the AP is modeled by the shared server and the dedicated DCH channels [12] of the NodeB are modeled by the dedicated servers. For simplicity, we refer to the several dedicated servers in [17] as one single dedicated server that consists of a pool of dedicated servers. Then the NodeB comprising dedicated DCH channels is modeled by this single dedicated server and this type of framework then fits well with our original setting in Section II. We thus preserve the two-server processing system as in Figure 1.

As mentioned earlier, the mobiles of dedicated streams directly join their respective AP or NodeB network. Mobiles arriving in the common stream decide to join one of the two networks based on their *estimate* of the expected service time in each one of them. However, an estimate of the expected service time of an arriving mobile j must be made taking into account the effect of subsequently arriving mobiles. But these subsequently arriving mobiles are themselves faced with a similar decision problem and hence their decision will affect the performance of mobile j (which is presently attempting to connect) or other mobiles already in service. This dependance thus induces a non-cooperative game structure to the decision problem and we seek here to study the Nash equilibrium solution of the game. The existence, uniqueness and structure of the equilibrium point have been proved in [17] already.

Here we seek to analytically determine the service time estimate (Section III-B) and explicitly compute the equilibrium achieving threshold policy (Section V). But before proceeding further in this direction, we briefly present below a background on the results of [17] adapted to our hybrid cell framework.

A. Background

A decision rule (or policy) for a new mobile is represented by a function $u : \{0, 1, \ldots, M_{AP}-1\} \to [0, 1]$, where M_{AP} is the capacity of the AP network. Thus for each possible state of the AP network denoted by number of mobiles already connected, m_c, a new mobile takes a randomized decision $u(m_c) \in [0, 1]$, that specifies the probability of connecting to the AP. $1 - u(m_c)$ then represents either the probability of connecting to the NodeB or abandoning to seek a connection altogether if both networks are full to their capacity. A policy profile $\pi = (u_1, u_2, u_3, \ldots)$ is defined as a collection of decision rules followed by all arriving mobiles indexed $1, 2, 3, \ldots$.

Define $V_{AP}(m_c, \pi)$ as the expected service time of a mobile in the AP network, given that it joins that network when m_c mobiles are already connected and all subsequently arriving mobiles follow the policy profile π. Mobiles in a NodeB network are allocated the *dedicated* DCH channels on which they are guaranteed throughputs greater than a worst case lower bound [12]. Equivalently, they are guaranteed service times lesser than a worst case upper bound. For simplification we assume a worst case estimate for the expected service time of a mobile in the NodeB network. Denote $\hat{\mu}_{3G} \triangleq \min_{\eta_j} \mu_{3G}(\eta_j)$ and let $\tau \triangleq 1/\hat{\mu}_{3G}$ be the maximum (or worst case) service time of a mobile j in the NodeB cell, which is *independent* of its load factor η_j.

Now, for some q ($q \in [0, 1]$, $q \in \mathbb{R}$), define the *decision* policy $u(m_c)$ to be the best response of a new mobile (against the policy profile $\pi = (u_1, u_2, u_3, \ldots)$ followed by all subsequently arriving mobiles), if,

$$u(m_c) = \begin{cases} 1 & : V_{AP}(m_c, \pi) < \tau \\ q & : V_{AP}(m_c, \pi) = \tau \\ 0 & : V_{AP}(m_c, \pi) > \tau \end{cases}$$

Further, define a special kind of policy namely the *threshold* type policy as follows. Given q and L such that $q \in [0, 1]$, $q \in \mathbb{R}$ and $L \in \mathbb{Z}^+ \cup \{0\}$, an L, q threshold policy $u_{L,q}$ is defined as,

$$u_{L,q}(m_c) = \begin{cases} 1 & : m_c < L \\ q & : m_c = L \\ 0 & : m_c > L \end{cases} \tag{3}$$

This L, q threshold policy will be denoted by $[L, q]$ or more compactly by $[g]$ where $g = L + q$. Note that the threshold policies $[L, 1]$ and $[L + 1, 0]$ are identical. We also use the notation $[g]^\infty \equiv [L, q]^\infty$ to denote the policy profile $\pi = ([g], [g], \ldots)$.

Now, it has been proved in Theorem 1 in [17] that a *Nash equilibrium achieving* decision policy (or simply an *equilibrium* policy), $u^*(m_c)$, exists, and it is actually the threshold policy, $[L^*, q^*]$, which can be computed as follows. If

$$V_{AP}(M_{AP} - 1, [M_{AP}]^\infty) < \tau,$$

then $[L^*, q^*] = [M_{AP}, 0]$. Otherwise, let

$$L^{min} \triangleq \min\{L \in \mathbb{Z}^+ \cup \{0\} : V_{AP}(L, [L, 1]^\infty) > \tau\}.$$

Now, if

$$V_{AP}(L^{min}, [L^{min}, 0]^\infty) \geq \tau,$$

then the threshold policy is given by $[L^*, q^*] = [L^{min}, 0]$. Else if

$$V_{AP}(L^{min}, [L^{min}, 0]^\infty) < \tau,$$

then it is given by $[L^*, q^*] = [L^{min}, q^*]$, where q^* is the unique solution of the equation,

$$V_{AP}(L^{min}, [L^{min}, q^*]^\infty) = \tau. \tag{4}$$

For explicitly computing the equilibrium policy $[L^*, q^*]$ we thus need to compute $V_{AP}(L, [L, 1]^\infty)$ for all possible values of L.

B. Determining Expected Service Time in AP

The entity equivalent to $V_{AP}(m_c, \pi)$ in [17] has been derived using *constant* departure rates. This scenario differs from our hybrid cell framework, since we consider a *state dependent* departure rate, $\mu_{AP}(m_c)$, for the shared AP server and moreover in our framework we have dedicated arrivals in addition to the common arrivals. Due to this difference we can not adopt the derivation of the entity equivalent to $V_{AP}(m_c, \pi)$ in [17]. Therefore, with the state dependent departure rate, $\mu_{AP}(m_c)$, we now compute $V_{AP}(m_c, \pi)$ analytically here.

For notational convenience if we define,

$$V(m_c) \triangleq V_{AP}(m_c, [L, q]^\infty), \quad 0 \leq m_c \leq M_{AP} - 1,$$

then the set of indeterminates $\{V(m_c) : 0 \leq m_c \leq M_{AP} - 1\}$ can be obtained as a solution of the following system of M_{AP} linear equations, where $\alpha \triangleq \lambda_{AP} + \lambda_{AP3G} + \mu_{AP}(m_c)$ (dependence of α on m_c has been suppressed in the notation):

Case 1: $4 \leq L \leq M_{AP} - 2$,

$$\begin{aligned}
V(0) &= \frac{1}{\alpha} + \frac{\lambda_{AP} + \lambda_{AP3G}}{\alpha} V(1) \\
V(m_c) &= \frac{1}{\alpha} + \frac{\mu_{AP}(m_c)}{\alpha} \frac{m_c}{m_c + 1} V(m_c - 1) \\
&\quad + \frac{\lambda_{AP} + \lambda_{AP3G}}{\alpha} V(m_c + 1), \quad 1 \leq m_c \leq L - 2 \\
V(L-1) &= \frac{1}{\alpha} + \frac{\mu_{AP}(L-1)}{\alpha} \frac{L-1}{L} V(L-2) \\
&\quad + \frac{\lambda_{AP} + q\, \lambda_{AP3G}}{\alpha} V(L) + \frac{\lambda_{AP3G}}{\alpha}(1-q) V(L-1) \\
V(L) &= \frac{1}{\lambda_{AP} + \mu_{AP}(L)} + \frac{\mu_{AP}(L)}{\lambda_{AP} + \mu_{AP}(L)} \frac{L}{L+1} V(L-1) \\
&\quad + \frac{\lambda_{AP}}{\lambda_{AP} + \mu_{AP}(L)} V(L+1) \\
V(m_c) &= \frac{1}{\lambda_{AP} + \mu_{AP}(m_c)} + \frac{\mu_{AP}(m_c)}{\lambda_{AP} + \mu_{AP}(m_c)} \frac{m_c}{m_c + 1} \\
&\quad \times V(m_c - 1) + \frac{\lambda_{AP}}{\lambda_{AP} + \mu_{AP}(m_c)} V(m_c + 1), \\
&\qquad L + 1 \leq m_c \leq M_{AP} - 2 \\
V(M_{AP} - 1) &= \frac{1}{\mu_{AP}(M_{AP} - 1)} + \frac{M_{AP} - 1}{M_{AP}} V(M_{AP} - 2)
\end{aligned} \tag{5}$$

Case 2: $L = M_{AP} - 1$,

$$\begin{aligned}
V(0) &= \frac{1}{\alpha} + \frac{\lambda_{AP} + \lambda_{AP3G}}{\alpha} V(1) \\
V(m_c) &= \frac{1}{\alpha} + \frac{\mu_{AP}(m_c)}{\alpha} \frac{m_c}{m_c + 1} V(m_c - 1) \\
&\quad + \frac{\lambda_{AP} + \lambda_{AP3G}}{\alpha} V(m_c + 1), \quad 1 \leq m_c \leq L - 2 \\
V(L-1) &= \frac{1}{\alpha} + \frac{\mu_{AP}(L-1)}{\alpha} \frac{L-1}{L} V(L-2) \\
&\quad + \frac{\lambda_{AP} + q\, \lambda_{AP3G}}{\alpha} V(L) + \frac{\lambda_{AP3G}}{\alpha}(1-q) V(L-1) \\
V(L) &= \frac{1}{\mu_{AP}(L)} + \frac{L}{L+1} V(L-1).
\end{aligned} \tag{6}$$

Each of the above equations says that the expected service time of a new mobile in AP cell ($V(m_c)$), when m_c other mobiles are connected, is given by the expected time till the next event (either arrival or departure) plus the expected service time from this event onwards. Note that $m_c = M_{AP}$ need not be considered since in that case the AP cell will be full to its capacity and can not accept a new mobile. The above system of M_{AP} linear equations in the framework of the non-cooperative game formulation is a very *generic* formulation of our user-network association decision problem and it can be solved using any particular WLAN and UMTS network models. In Section V, we will solve this problem assuming a specific definition for the packet throughputs (or departure rates) obtained from specific models for the WLAN and UMTS networks. We first present these network models in the following section along with some assumptions.

IV. WLAN and UMTS Network Models

Before discussing the network models adopted from previous work, we first state below some assumptions along with their justification. Since the bulk of data transfer for a mobile engaged in streaming or interactive (HTTP like) data transmission is carried over the downlink (AP to mobile or NodeB to mobile) and since TCP is the most commonly used transport protocol (streaming protocols based on TCP also exist, e.g., Real Time Streaming Protocol), we are interested here in network models for computing TCP throughput on only downlink.

A. Assumptions

1) Assumption on QoS and TCP: We assume a single QoS class of arriving mobiles so that each mobile has an identical minimum downlink throughput requirement of θ_{min}, i.e., each arriving mobile must achieve a downlink packet throughput of at least θ_{min} bps in either of the two networks.

Several versions of TCP have been proposed in literature for wireless environments. For our purposes we assume that the wireless TCP algorithm operates in *split mode* [16]. In brief, the split mode divides a TCP connection into wireless and wired portions, and acks are generated for both portions separately. Therefore, in our hybrid cell scenario TCP acks are generated separately for the single hop between mobiles and AP or NodeB. We also assume that TCP acks are not delayed and every received data packet is acknowledged with an ack. It is further assumed that each mobile's or receiver's advertised window W^* is set to 1 in the wireless portion of TCP protocol.

2) Resource allocation in AP and NodeB: We assume *saturated resource allocation* in the downlink of AP and NodeB networks. Specifically, this assumption for the AP network means that the AP is *saturated* and has infinitely many packets backlogged in its transmission buffer. In other words, there is always a packet in the AP's transmission buffer waiting to be transmitted to each of the connected mobiles. With this assumption, mobiles can be allocated downlink packet throughputs greater than their QoS requirements of θ_{min} and cell resources in terms of transmission opportunities (*TxOPs*) on the downlink will be maximally utilized.

In the NodeB network the saturated resource allocation assumption implies that at any given instant the NodeB cell resources on downlink are fully utilized. This is analogous to the maximal utilization of *TxOPs* in the AP network discussed in previous paragraph. With this maximum resource allocation assumption even if a mobile has a minimum packet throughput requirement of only θ_{min} bps, it can actually be allocated a higher throughput if additional unutilized cell resources are available.

3) Justification: The assumption of W^* being set to 1 is required for the WLAN model that we adopt and in fact it is known to provide the best performance of TCP in a single hop case (see [10], [11] and references therein).

Saturated resource allocation is a standard assumption, usually adopted to simplify modeling of complex network frameworks like those of WLAN and UMTS (see for e.g., [12], [13]).

B. Downlink Throughput in 802.11 WLAN AP

We reuse the downlink TCP throughput formula for a mobile in an 802.11 WLAN from [15]. For completeness, here we briefly mention the network model that has been extensively studied in [15] and then simply restate the throughput expression without going into much details.

Each mobile connected to the AP uses the Distributed Coordination Function (DCF) protocol with an RTS/CTS frame exchange before any data-ack frame exchange and each mobile (including the AP) has an equal probability of the channel being allocated to it. The AP does not employ any rate control algorithm and transmits at a fixed PHY data rate of R_{data} bps to all mobiles. With the assumption of W^* being set to 1 (Section IV-A), any mobile will always have a TCP ack waiting to be sent back to the AP with probability $1/2$, which is also the probability that it contends for the channel. This is however true only for those versions of TCP that do not use delayed acks. If the AP is always saturated or backlogged, the average number of backlogged mobiles contending for the channel is given by $m_b = 1 + \frac{m_c}{2}$. Based on this assumption and since for any connection an ack is sent by the mobile for every TCP packet received, the downlink TCP throughput of a single mobile is given by Section 3.2 in [15] as,

$$\theta_{AP}(m_c) = \frac{L_{TCP}}{m_c(T_{TCPdata} + T_{TCPack} + 2T_{tbo} + 2T_w)}, \tag{7}$$

where L_{TCP} is the size of TCP packets and $T_{TCPdata}$ and T_{TCPack} are the raw transmission times of a TCP data and a TCP ack packet, respectively. T_{tbo} and T_w denote the mean total time spent in *back-off* and the average total time wasted in collisions for any successful packet transmission and are computed assuming m_b backlogged mobiles. The explicit expressions for $T_{TCPdata}$, T_{TCPack}, T_{tbo} and T_w can be referred to in [15]. However, we mention here that they depend on certain quantities whose numerical values have been provided in Section V. Note that independent of their location in the AP cell, all mobiles achieve equal downlink TCP throughputs (given by Equation 7) in a fair manner [15].

C. Downlink Throughput in 3G UMTS NodeB

We consider a standard model for data transmission on downlink in a 3G UMTS NodeB cell. Let W be the WCDMA modulation bandwidth and if $SINR_j$ denotes the signal to interference plus noise ratio received at a mobile j then its energy per bit to noise density ratio is given by,

$$\left(\frac{E_b}{N_o}\right)_j = \frac{W}{\theta_{3G}} \times SINR_j. \tag{8}$$

From Chapter 8 in [12] we can then say that in a NodeB cell with saturated resource allocation, the downlink TCP throughput, θ_{3G}, of any mobile j as a function of its load factor η_j is given by,

$$\theta_{3G}(\eta_j) = \frac{\eta_j\, W}{(E_b/N_o)_j\,(1-\alpha_j+i_j)}, \tag{9}$$

where i_j and α_j are mobile j's inter-cell to intra-cell interference ratio and orthogonality factor, respectively (see Section 8.2.2.2 in [12]).

It is to be noted here that the required $(E_b/N_o)_j$ ratio by each mobile j is a function of its throughput. Also, if each mobile operates at its minimum throughput requirement of θ_{min} then we can easily compute the capacity, M_{3G}, of the cell as,

$$M_{3G} = \sum_j \frac{\eta_j\, W}{\theta_{min}\,(E_b/N_o)_j\,(1-\alpha_j+i_j)}, \tag{10}$$

where, the summation is over all mobiles in the NodeB cell. Note that the load factor, η_j, of a mobile j decreases with increasing number of total mobiles in a NodeB cell [12].

For some values of η_j in the interval $[0.09, 0.9]$ and $\theta_{min} =$ 115 kbps, Table I shows the SINR (second column) received at any mobile j as a function of its load factor (first column). Note that we consider a maximum load factor of 0.9 and not 1 in order to avoid instability conditions in the cell. These values of SINR have been obtained from radio layer simulations of a NodeB cell. The third column shows the downlink packet throughput with a block error rate (BLER) of 10^{-2} that can be achieved by a mobile as a function of the SINR observed at that mobile. And the fourth column lists the corresponding values of $(E_b/N_o)_j$ ratio (obtained from Equation 8) that are required at mobile j to successfully decode NodeB's transmission.

η_j	$SINR_j$ (dB)	$\theta_{3G}(\eta_j)$ $(kbps)$	$\left(\frac{E_b}{N_o}\right)_j$ (dB)
0.9	0.8423	572	9.0612
0.45	−2.1804	465	6.9503
0.3	−3.7341	405	5.7894
0.225	−5.1034	360	5.0515
0.18	−6.0327	322	4.5669
0.15	−6.5093	285	4.3052
0.1286	−7.2075	242	4.3460
0.1125	−8.8312	191	4.7939
0.1	−8.9641	144	5.5091
0.09	−9.1832	115	6.0281

TABLE I

V. Computing the Equilibrium Policy

With the network models defined in previous section, we now solve the system of M_{AP} linear equations in order to obtain the set of indeterminates $\{V(m_c) : 0 \le m_c \le M_{AP}-1\}$ and finally the equilibrium threshold policy $[L^*, q^*]$, as described in Section III. The focus of our numerical analysis here is to study the equilibrium policy under an ordinary network scenario. We do not investigate in detail the effects of specific TCP parameters and it is outside the scope of this paper. The network scenario that we consider is as follows: $L_{TCP} = 8000$ bits (size of TCP packets), $L_{MAC} = 272$ bits, $L_{IPH} = 320$ bits (size of MAC and TCP/IP headers), $L_{ACK} = 112$ bits (size of MAC layer ACK), $L_{RTS} = 180$ bits, $L_{CTS} = 112$ bits (size of RTS and CTS frames), $R_{data} = 11$ Mbits/s, $R_{control} = 2$ Mbits/s (802.11 PHY data transmission and control rates), $CW_{min} = 32$ (minimum 802.11 contention window), $T_P = 144\mu s$, $T_{PHY} = 48\mu s$ (times to transmit the PLCP preamble and PHY layer header), $T_{DIFS} = 50\mu s$, $T_{SIFS} = 10\mu s$ (distributed inter-frame spacing time and short inter-frame spacing time), $T_{slot} = 20\mu s$ (slot size time), $K = 7$ (*retry limit* in 802.11 standard), $b_0 = 16$ (initial mean back-off), $p = 2$ (exponential back-off multiplier), $\lambda_{AP} = 3$, $\lambda_{3G} = 3$, $\lambda_{AP3G} = 10$, $1/\zeta = 10^5$ bits, $\alpha_j \in [0.6, 0.9]$ uniformly randomly [12], $i_j \in [0.4, 0.7]$ uniformly randomly [12] and $W = 3.84$ Mcps.

Plugging Equation 7 in Equation 1 and then Equation 1 in the system of M_{AP} linear equations, it can be solved with $m_c = L$ and $q = 1$ to obtain $V_{AP}(L, [L, 1]^\infty)$ for different values of L. Figure 2 shows an example plot of $V_{AP}(L, [L, 1]^\infty)$ for $1/\zeta = 10^5$ bits, $\lambda_{AP} = 3$, $M_{AP} = 10$ and other numerical values for various entities in WLAN and UMTS networks being those listed in the previous paragraph.

Assuming a certain NodeB cell capacity, M_{3G}, τ can be computed from its definition, Equation 10 and Table I. Knowing τ, one can compute L^{min} from Figure 2. It is simply the smallest integer value of L for which $V_{AP}(L, [L, 1]^\infty) > \tau$. Using this value of L^{min} one can finally compute q^* from Equation 4 which will give us the equilibrium threshold policy.

As an example, a capacity of $M_{3G} = 10$ (i.e., $\theta_{min} =$ 115 kbps from Equation 10 and Table I) and $\lambda_{AP} = 3$ gives rise to a value of $\tau = 2.5$. For $\tau = 2.5$, we obtain different values of L^{min} for different values of λ_{AP3G} from Figure 2. Finally, these values of L^{min} are then used to compute q^* from Equation 4 and Figure 3 shows a plot of the equilibrium

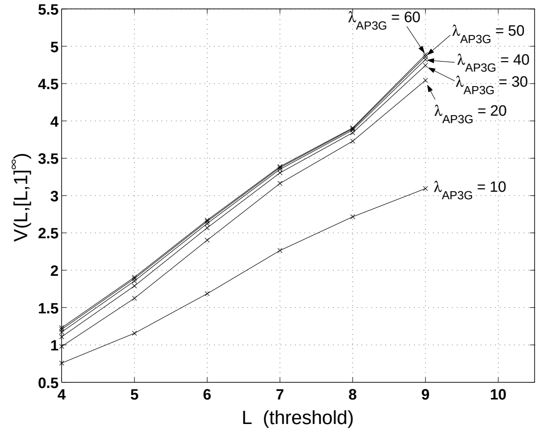

Fig. 2. $V_{AP}(L, [L, 1]^{\infty})$ v/s L for $\lambda_{AP} = 3$ and $M_{AP} = 10$

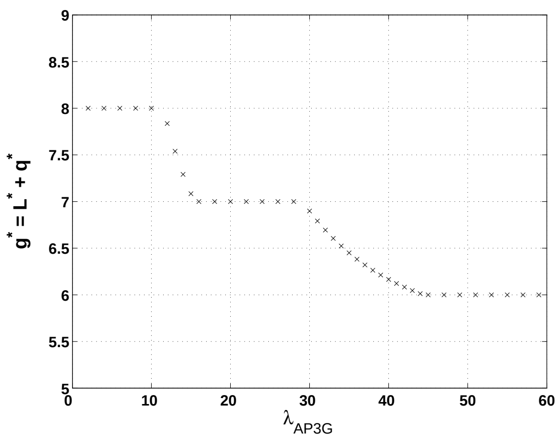

Fig. 3. g^* v/s λ_{AP3G} for $\lambda_{AP} = 3$, $M_{AP} = 10$ and $\tau = 2.5$

threshold, $g^* = L^* + q^* = L^{min} + q^*$, against various values of λ_{AP3G}. In this figure a value of $g^* = 7.34$ implies that $L^* = 7$ and $q^* = 0.34$. We clearly observe here that the equilibrium threshold has a special structure of *descending staircase* with increasing arrival rate (λ_{AP3G}) of mobiles in common stream. This special structure is due to the way the threshold type policy has been defined in Equation 3. As the value of λ_{AP3G} increases the equilibrium threshold decreases. This implies that for high values of λ_{AP3G} (> 44), it is (individually) optimal for mobiles of common stream to join the AP network (w.p. 1) only if there are less than 6 mobiles already connected to AP. Otherwise, it is (individually) optimal for them to connect to the NodeB network (see Equation 3). For low values of λ_{AP3G} (< 28), it is (individually) optimal for common stream mobiles to join the AP even if there are 7 or 8 (or less) mobiles already connected to AP. Recall that the AP cell capacity considered in this example is $M_{AP} = 10$. For certain values of λ_{AP3G} we get a non-integer threshold. This implies that (individually) optimal performance (in terms of average service time of each mobile) is achieved when the common stream mobiles connect to AP with a certain probability (q^*, or q in Equation 3) and to NodeB otherwise.

VI. CONCLUSION

In this paper, we have considered individually optimal user-network association in an AP-NodeB hybrid cell. To the best of our knowledge this study is the first of its kind. Since it is infeasible to obtain the equilibrium policy for an exhaustive set of network scenarios, we have considered here an ordinary network scenario and explicitly computed the equilibrium policy. Even though the characteristics of the solution to our particular scenario are not depictive of the complete solution space, they can certainly be helpful in acquiring an intuition about the underlying dynamics of individually optimal user-network association in a hybrid cell.

REFERENCES

[1] L. Ma, F. Yu, V. C. M. Leung, T. Randhawa, *A New Method to support UMTS/WLAN Vertical Handover using SCTP*, IEEE Wireless Commun., Vol.11, No.4, p.44-51, Aug. 2004.

[2] J. Song, S. Lee, D. Cho, *Hybrid Coupling Scheme for UMTS and Wireless LAN Interworking*, Proc. VTC-Fall, Vol.4, pp.2247-2251, Oct. 2003.

[3] C. Liu, C. Zhou, *HCRAS: A Novel Hybrid Internetworking Architecture between WLAN and UMTS Cellular Networks*, IEEE CCNC, Las Vegas, Jan. 2005.

[4] M. Jaseemuddin, *An Architecture for integrating UMTS and 802.11 WLAN Networks*, Proc. ISCC 2003, Turkey, July 2003.

[5] N. Vulic, S. H. Groot, I. Niemegeers, *Common Radio Resource Management for WLAN-UMTS Integration Radio Access Level*, Proc. IST Mobile & Wireless Communications Summit 2005, Germany, June 2005.

[6] H. Kwon, K. Rho, A. Park, J. Ryou, *Mobility Management for UMTS-WLAN Seamless Handover: within the Framework of Subscriber Authentication* Proc. ISATED Communication, Network and Information Security (CNIS), Nov. 2005.

[7] O. E. Falowo, H. A. Chan, *AAA and Mobility Management in UMTS-WLAN Interworking* Proc. 12th International Conference on Telecommications (ICT), Cape Town, May 2005.

[8] K. Premkumar, A. Kumar, *Optimal Association of Mobile Wireless Devices with a WLAN-3G Access Network*, Proc. IEEE ICC, June 2006.

[9] F. Yu, V. Krishnamurthy, *Efficient Radio Resource Management in Integrated WLAN/CDMA Mobile Networks*, Telecommunication Systems Journal, Vol.30, No.1-3, p.177-192, Nov. 2005.

[10] Fu et al., *The Impact of Multihop Wireless Channel on TCP Throughput and Loss*, Proc. IEEE INFOCOM, 2003.

[11] F. Lebeugle, A. Proutiere, *User-level performance in WLAN hotspots*, Proc. ITC, Beijing, 2005.

[12] H. Holma, A. Toskala, *WCDMA for UMTS*, Revised Edition, Wiley, 2001.

[13] A. Kumar, E. Altman, D. Miorandi and M. Goyal, *New insights from a fixed point analysis of single cell IEEE 802.11 WLANs*, Proc. IEEE Infocom, USA, March 2005.

[14] T. Bonald, *A score-based opportunistic scheduler for fading radio channels*, Proc. European Wireless, 2004.

[15] D. Miorandi, A. Kherani, E. Altman, *A Queueing Model for HTTP Traffic over IEEE 802.11 WLANs*, Computer Networks, Vol.50, Issue 1, p.63-79, Jan'06.

[16] Y. Tian, K. Xu, N. Ansari, *TCP in Wireless Environments: Problems and Solutions*, IEEE (Radio) Communications Magazine, Vol. 43, No. 3, pp. S27-S32, March 2005.

[17] E. Altman, N.Shimkin, *Individual equilibrium and learning in a processor sharing system*, Operations Research, Vol.46, p.776-784, 1998.