On the Impact of Soft Vertical Handoff on Optimal Voice Admission Control in PCF-Based WLANs Loosely Coupled to 3G Networks

Racha Ben Ali, Student Member, IEEE, and Samuel Pierre, Senior Member, IEEE

Abstract—Soft vertical handoff (VHO) and admission control are usually considered as two independent mechanisms ensuring respectively packet-level QoS and call-level QoS for voice calls in loosely coupled 3G/WLAN networks. In this paper, we evaluate the impact of the soft VHO on the blocking performance of the optimal voice admission control in different mobility environments where the WLAN operates the Point Coordination Function (PCF). For this purpose, we propose an accurate analytical mobility model for the soft VHO region. Then, based on the proposed model, we derive and analyze the blocking and dropping probability expressions of the optimal voice admission control algorithms in the 3G network loosely coupled to the PCF-based WLAN. Results show us that a resource-efficient soft handoff (RESHO) performs significantly better than a static-threshold soft handoff (STSHO) particularly in WLAN mobility environments. In fact, the 3G new call blocking probability reduction gained by using RESHO compared to STSHO is largely increased when mobile station (MS) velocities have low mean and high variability which typically characterizes the WLAN mobility environment. Besides, results show us that RESHO reduces all blocking and dropping probabilities. We believe that the provided model and the presented results could help design efficient MS controlled soft VHO algorithms for emergent loosely coupled 3G/WLAN networks.

Index Terms—Loosely coupled UMTS/WLAN, soft vertical handoff, mobility model, voice optimal CAC, blocking probability.

I. INTRODUCTION

Wireless local area network (WLAN) technology have been widely deployed in recent years and is still experiencing a tremendous growth due to its flexibility, high-rate access and low cost. On the other hand, voice over Internet protocol (VoIP) has become one of the fastest growing Internet applications because of its high bandwidth efficiency and low cost. In the near future, voice over WLAN (VoWLAN) which can be viewed as the convergence of these two technologies is expected to have a dramatic popularity. However, voice on today’s popular WLAN systems still suffers from several quality of service (QoS) issues due to its real-time constraints which make it an active research area. In fact, a lot of recent research [1]–[3] has been carried out in order to resolve packet-level QoS issues of VoWLAN by providing efficient admission control mechanisms and fast handoff procedures. Particularly, the primary operation mode of WLAN called the Distributed Coordination Function (DCF) makes the voice traffic; which is very sensitive to delay and jitter; interfere with other types of traffic resulting in a violation of its delay bound or in a large delay variance [4]. In our work, we consider a voice-centric WLAN provider implementing the optimal Point Coordination Function (PCF) which, as opposed to the very popular DCF, provides strict voice QoS by designing a simple admission control based on [5]. Note that PCF also provides better security compared to DCF since denial of service attacks by outside users can be better controlled by a PCF based WLAN.

Dealing with VoWLAN QoS issues related to user mobility, several WLAN smart scanning and fast handoff algorithms minimizing the service disruption during handovers have been presented in [3]. For instance, authors in [6] used a neighbor graph cache mechanism to reduce scanning latency down to a level that meets the voice QoS delay constraints during WLAN handoff. However, even though recent proposed solutions can provide suitable fast handoff for VoWLAN users as stated in [3], the small coverage of WLAN networks is a major limitation for providing the required ubiquity to mobile voice over IP users. Hopefully this limitation can be bypassed by integrating the WLAN with a large coverage overlaying 3G cellular network. In fact, with the increasing popularity of multi-mode mobile stations (MS) equipped with both cellular and WLAN interfaces, such as recent PDAs and smartphones, voice users will profit from both the low cost of high-speed WLAN access in hotspot areas and the ubiquitous coverage of 3G cellular networks. For this purpose, we assume that 3G cellular networks and WLAN networks are integrated in order to provide voice call continuity. Furthermore, considering the benefits of implementation flexibility and independent deployment of 3G and WLAN networks, we assume a loose coupling scheme that integrates these heterogeneous wireless networks using an external IP-based network.

However, it is known that an IP-based loosely coupled integration induces high handoff latencies especially for handoffs involving IP address and location updates such as in vertical handoffs using either SCTP or Mobile IP protocols. Therefore, a soft handoff mechanism is needed in order to ensure the packet-level QoS of voice calls during the disruption period.
of the handoff process. In general, the soft handoff provides simultaneous connectivity to more than one base station (BS) / access point (AP) in order to seamlessly transfer the call between them. Considering the introduction of the soft handoff mechanism in these integrated networks, standard optimal admission control mechanisms have to be adapted in order to provide efficient wireless bandwidth utilization while maintaining the required packet-level and call-level QoS for the voice service. In the literature, efficient call admission control (CAC) in loosely coupled 3G/WLAN networks is analyzed in [7] and soft handoff is introduced in [8], [9]. However, to the best of our knowledge no research work was conducted to evaluate the impact of the soft handoff on the voice call blocking performance of optimal admission control.

The rest of this paper is organized as follows. In Section II, we propose our advanced analytical model taking into account both mobility and soft handoff characteristics in 3G/WLAN integrated networks. Then in Section III, we present the resource-efficient soft handoff algorithm for ensuring the packet-level QoS and we analyze its related parameters. In Section IV, we provide our adaptation of the optimal admission control algorithm to a 3G/WLAN integrated network in order to ensure the call-level QoS by maximizing the number of accepted new calls while maintaining a hard constraint on a very low call dropping rate. In Section V, we validate the analytical model using computer simulation and we present the blocking performance results of soft vertical handoff algorithms under different mobility schemes. Finally, Section VI concludes the paper.

II. SYSTEM MODEL

In our system model, we consider a loosely coupled 3G/WLAN system in which several WLAN cells overlay each 3G cell. Within this system, location dependant mobility and traffic distribution are adequately modeled, voice capacity analysis is briefly discussed, blocking probabilities are estimated and soft handoff parameters are analyzed. Since many symbols are used, Table I summarizes the important ones.

A. Multi-region mobility model

In general, mobility can be modeled using either analytical or simulation models. Simulation is able to provide realistic mobility models since it uses a large amount of detail to build the trajectory of the mobile user by periodically tracking its location in small time steps. However such models are generally intractable when considering large network coverage topologies with large number of users. Therefore, analytical mobility models based on random variables and stochastic processes used to model cell residence time are more suitable in this case. In our work we develop a new analytical mobility model in 3G/WLAN coverage to study the impact of soft handoff on resource utilization. We propose a location dependant mobility model within a 3G cell which is composed of three types of regions based on the number of overlaying coverages (single or double) and the number of occupied resources (uni-casting or bi-casting) illustrated in Figure 2: several WLAN core regions, their correspondent WLAN border regions designed for soft handoff and finally one single coverage 3G region. These regions are defined as following:

- A WLAN core region, noted RW, is characterized by the double coverage of both WLAN and 3G networks but only resources of one coverage are allocated at a time to a multi-mode MS. We assume a network selection strategy that gives resource allocation preference to the low-cost WLAN rather than the generally highly loaded 3G. Due to the low mobility in WLAN environments (indoor environments in general) most users stay within this region for a short time, while few users stay for a very long time. It was shown in [10] that heavy-tailed Pareto distributions provide the best fit to captured realistic WLAN residence times. Besides, it was shown that usual exponential distributions are no more valid for modeling WLAN residence time. Therefore in order to capture the Pareto effect without loosing the Markov property needed in queuing analysis, we use a hyper-exponential distribution with a mean rate \( \eta_w \) and a mobility variability parameter \( \alpha_w \) to model the residence time in this RW region illustrated in Figure 1. Thus, the probability density function (pdf) of this residence time is expressed as following:

\[
f_{\text{RW}}(t) = \frac{\alpha_w}{\alpha_w^\eta_w} e^{-\alpha_w^\eta_w t} \left( \frac{1}{\eta_w} + \frac{1}{\alpha_w^\eta_w} \right) e^{-\frac{\eta_w}{\alpha_w^\eta_w} t} \]  

- A soft handoff WLAN border region, noted RWC (resp. RCW), is characterized by the double coverage of both WLAN and 3G networks and in which duplicate resources are allocated on both networks at the same time to the multi-mode MS moving out of (resp. into) the WLAN core region. Particularly, it defines the soft vertical handoff (VHO) region in which the multi-mode MS maintains two simultaneous connections by bi-casting the ongoing voice call on both WLAN AP and 3G BS. This bi-casting in this WLAN border region provides the required packet-level QoS for voice calls during the handoff disruption period. The call-level QoS can be significantly improved by recovering call droppings induced by WLAN horizontal handoff (HHO) failures using the soft upward VHO, i.e. the soft handoff that overflows the call from WLAN to 3G. Depending on the mobility management architecture implemented by 3G and WLAN operators, the soft handoff may involve Mobile IP bi-casting or SCTP multi-homing mechanisms. Since we are interested in the handoff part that involves the duplicate resources defining the soft handoff, the handoff delay represents the delay of the wireless signalling process immediately after the channel reservation on the next AP/BS added to the delay of the IP address configuration and the location update on the correspondent node (CN). We model this handoff delay by the sum of a constant added to a non negative random variable that depends on IP address configuration, CN location and Internet latency. Since a constant distribution does not keep the Markovian property [11] required for our further queuing analysis, we approximate it using the Erlang distribution with a large number of \( r \) stages. Therefore, we model the total residence time in a soft VHO region using the sum of \( r \)
TABLE I
SUMMARY OF IMPORTANT SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>Index representing a WLAN cell if $k=w$ and a 3G cell if $k=c$</td>
</tr>
<tr>
<td>$\lambda_k$</td>
<td>Mean new call arrival rate to cell $k$</td>
</tr>
<tr>
<td>$\mu_k$</td>
<td>Mean call holding time</td>
</tr>
<tr>
<td>$\mu_k^{RW}$ (resp. $\mu_k^{w}$)</td>
<td>Mean channel holding time for a new (handoff) call in cell $k$</td>
</tr>
<tr>
<td>$\eta_k$</td>
<td>WLAN (3G cell) core region residence time mean parameter</td>
</tr>
<tr>
<td>$\alpha_k$ (resp. $\alpha_w$)</td>
<td>WLAN (3G cell) core region residence time variability parameter</td>
</tr>
<tr>
<td>$\sigma_k$ (resp. $\sigma_w$)</td>
<td>Mean of the variable part (constant part) of the residence time in the cell $k$ soft handoff border region</td>
</tr>
<tr>
<td>$\eta_{cw}$</td>
<td>WLAN residence time mean parameter</td>
</tr>
<tr>
<td>$P(R_{WT})$ (resp. $P(R_{Wc})$)</td>
<td>WLAN (3G cell) residence time</td>
</tr>
<tr>
<td>$P^0_{f}$ (resp. $P^0_{c}$)</td>
<td>Probability of a user moving out of current 3G cell to overlaying WLAN cell</td>
</tr>
<tr>
<td>$v_i$ (resp. $v_w$)</td>
<td>Vertical (Horizontal) handoff rate in (to) cell $k$</td>
</tr>
</tbody>
</table>

CAC parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_s$ (resp. $g_w$)</td>
<td>Limited fractional guard channel LFGC for new (vertical) handoff calls in (to) cell $k$</td>
</tr>
<tr>
<td>$f_k(\ldots)$ (resp. $f_w(\ldots)$)</td>
<td>Probability acceptance function for new calls (vertical handoff calls) in cell $k$</td>
</tr>
<tr>
<td>$p_{d,max}^k$ (resp. $p_{d,max}^w$)</td>
<td>(Maximum allowed) Call dropping probability for calls in the separated $k$ network</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Resolution of the estimated handoff rates and steady-state probabilities</td>
</tr>
</tbody>
</table>

QoS metrics

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^0_{b}$ (resp. $P^0_{c}$)</td>
<td>Probability of blocking new calls requesting admission to the $k$ network</td>
</tr>
<tr>
<td>$P_{f,h}$</td>
<td>Probability of failing the horizontal (vertical) handoff in (to) the $k$ network</td>
</tr>
<tr>
<td>$P_{f,h}$ (resp. $P_{f,w}$)</td>
<td>Probability that a new (handoff) call hand off from the cell $k$</td>
</tr>
<tr>
<td>$P_{f,h}$ (resp. $P_{f,w}$)</td>
<td>Probability that a call (either new call or handoff call) hand off from the cell $k$</td>
</tr>
</tbody>
</table>

Received signal strength parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{SS}$</td>
<td>Received signal strength measured at time 0s</td>
</tr>
<tr>
<td>$R_{SS,s}$</td>
<td>Received signal strength threshold that triggers the move into the WLAN cell after moving out from it</td>
</tr>
<tr>
<td>$R_{SS,s}$ (resp. $R_{SS,s}^{w}$)</td>
<td>Received signal strength threshold that triggers the static-threshold soft handoff to neighboring WLAN cell or 3G overlaying cell</td>
</tr>
<tr>
<td>$R_{SS,m}$</td>
<td>Received signal strength threshold that triggers the move out from the WLAN cell</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Slope index of the received signal strength model</td>
</tr>
<tr>
<td>$d(.)$</td>
<td>Distance between the MS and the AP function of time</td>
</tr>
<tr>
<td>$v_i$</td>
<td>Velocity (Angle between WLAN cell diameter and movement direction) of the MS in the WLAN during time interval $i$</td>
</tr>
<tr>
<td>$\tau_{id}$</td>
<td>Ideal time interval in which the resource-efficient soft handoff algorithm trigger itself</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Time interval duration for averaging the RSS and estimating the RSS slope under current velocity</td>
</tr>
</tbody>
</table>

Other parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Number of WLAN cells overlaying one single 3G cell</td>
</tr>
<tr>
<td>$r$</td>
<td>Number of stages of the Erlang distribution used to approximate a constant distribution</td>
</tr>
<tr>
<td>$P_{new}$</td>
<td>Probability that no other AP can be discovered before reaching the out of range RSS on the current AP.</td>
</tr>
</tbody>
</table>

Erlang distributions with rate $K_w$ for soft upward VHO (resp. $K_c$ for soft downward VHO) and an exponential distribution with a mean rate $\sigma_w$ (resp. $\sigma_c$). $r_{SS}/K_w$ is the constant part of the handoff delay and $1/\sigma_w$ is the mean parameter of its variable part. In a further section, we will distinguish between a theoretical resource-efficient and a practical less resource-efficient soft handoff regions. The first one is optimally dimensioned for the strict latency of the handoff procedure with a parameter $1/\sigma_w^*$. The second one is dimensioned for a higher value $1/\sigma_w$. The RWC residence time pdf is expressed as follows:

$$f_{RWC}(t) = (t - K_w)e^{-\sigma_w(t-K_w)}$$

(2)

- A 3G cell core region, noted RC, is characterized by a single 3G coverage. The residence time in this region is modeled, as in RW, using a hyper-exponential distribution with a mean rate $\eta_c$, and a mobility variability parameter $\alpha_c$. We use this distribution to overcome the no more valid exponentially distributed CRT assumption and to capture the CRT high variability in next generation cellular systems as noticed in [12]. Note that the RC residence time pdf is similar to the RW one.

B. Voice capacity model

Depending on the access technology used, there are several analytical methods in the literature that bound the maximal number of voice calls, i.e. the number of channels, that a cell either 3G or WLAN can accommodate with the required packet-level QoS.
1) **3G cell voice capacity model:** We assume that the 3G network is based on the popular UMTS standard which uses the wide band code-division multiple access (W-CDMA) cellular system. Assuming different types of calls, the W-CDMA soft capacity is modeled using the very popular load expression in [13]. Using the expression in [13] that bounds the interference, we can estimate the 3G cell voice capacity in term of maximal number of voice calls which is a function of the actual number of calls admitted in each other class. In order to simplify our analysis we assume that voice calls are preemptively prioritized over other types of calls. Therefore, the voice capacity region can be easily evaluated independently from other types of calls.

2) **WLAN cell voice capacity model:** Even though the DCF-based QoS differentiation mechanism introduced in the recent IEEE 802.11e standard [14] can provide transmission priority to voice traffic it can not guarantee its delay bound since it cannot preempt ongoing transmissions of other low-priority traffic using the distributed operation mode. However, despite the quite lower resource efficiency of the PCF operation mode when no voice traffic is present due to the unnecessary MS polling, this centralized access method can provide strict delay guarantees for voice calls, as opposed to the DCF. The algorithm proposed in [5] provides an analytical expression giving the number of nodes that a PCF-based AP can accommodate in order to satisfy a given delay constraint at each of the nodes. In the paper, it was shown that an AP configured with Contention Free Period repetition interval $CFP_{\text{pri}}$ of 25 ms can ensure the strict QoS of 18 voice calls using the G.723.1 codec. In our work, it is possible to limit the maximal number of voice calls to 15 so that a minimal Contention Period (CP) will be guaranteed for data traffic using the DCF as an example of a restricted access resource sharing scheme [15]. In fact, if the MS has no active voice sessions it will be removed from the polling list and a longer CP will be reserved for data traffic.

C. The admission control policy

Usually when we can derive the voice capacity $c_v$ for the 3G cell [13] (resp. $c_w$ for the PCF-based WLAN cell [5]) in terms of maximal number of voice calls supported without violating the packet-level QoS, a simple admission control algorithm can be designed. However, since it is commonly considered that the dropping of an ongoing call has much more negative impact on users' perception than the blocking of a newly initiated call, we have to prioritize handoff calls over new ones in accessing the wireless bandwidth resources. An acceptance probability function $\beta_k(i,j)$ (resp. $\check{\beta}_k(i,j)$) similar to the one used in [16] is designed for admitting newly initiated voice calls (resp. VHO calls) in the loosely coupled 3G/WLAN network. Where $k$ is either $w$ for WLAN or $c$ for cellular 3G. We assume that since in a loosely coupled 3G/WLAN network the 3G and the WLAN bandwidth resources belong usually to two distinct administrative domains, VHO calls which are controlled and triggered at the MS level are processed as new calls. And therefore the priority to access the voice capacity for VHO calls and for new calls are equals ($\check{\beta}_k(i,j) = \beta_k(i,j)$). This acceptance function is tightly related to the limited fractional guard channel (LFGC) [16] which defines the non-integer number of channels $g_k$ exclusively reserved for HHO calls. Up to $c_k - g_k$ channels can be occupied by new calls and up to $c_k - g'_k$ can be occupied by VHO calls. Note that $g'_k = g_k$ since $\check{\beta}_k(i,j) = \beta_k(i,j)$. In fact, $\beta_k(i,j)$ is equal to 1 if $i + j < [c_k - g_k]$, $[g_k] - g_k$ if $i + j = [c_k - g_k]$ and 0 if $i + j > [c_k - g_k]$. $g_k$ have to be chosen as the optimal solution of the CAC problem presented further.

D. Expression of channel holding times (CHT)

Let $\lambda_w$ (resp. $\lambda_c$) denote the new call arrival rate to a WLAN cell (resp. to a 3G cell) and let $1/\mu$ denote the mean voice call holding time, i.e. its mean duration. It is commonly assumed that new calls arrive following a Poisson process and that the call holding time is exponentially distributed. For sake of notation simplicity, we note $WRT$ (resp. $CRT$) the random variable representing the WLAN cell residence time (resp. the 3G cell residence time). Besides, we note $RWT$, $RCW$ and $RCWT$ the random variables representing the residence times in their respective regions. Since $WRT$ is the sum of two non-negative independent continuous random variables $RW$ and $RCW$, i.e. $WRT = RW + RCW$, its pdf is the convolution of the pdf of $RW$ and the pdf of $RCW$:

$$f_{WRT}(t) = (f_{RW} \ast f_{RCW})(t) = \int_0^t f_{RW}(x)f_{RCW}(t-x)dx$$

The Laplace transform of $WRT$’s pdf is easier to develop since it is the product of Laplace transforms of $RW$’s pdf and $RCW$’s pdf. Thus applying the Laplace transform to Equations 1 and 2 we obtain:

$$F_{RW}(s) = \frac{(c^2_w - \alpha_w + 1)s + \alpha_w \eta_w}{(\alpha_w s + \eta_w)(s + \alpha_w \eta_w)}$$

$$F_{RCW}(s) = \frac{K_w}{K_w + s}$$

$$F_{WRT}(s) = F_{RW}(s)F_{RCW}(s)$$

$$= \frac{(c^2_w - \alpha_w + 1)s + \alpha_w \eta_w}{(\alpha_w s + \eta_w)(s + \alpha_w \eta_w)} \frac{K_w}{K_w + s}$$

Assuming a channelized PCF-based WLAN bandwidth for voice calls as previously described and a conservative maximal number of voice calls $ck$ in the loosely coupled 3G/WLAN network and let $1/\mu$ denote the mean voice call holding time, i.e. its mean duration. Since $WRT$ is the sum of two non-negative independent continuous random variables $RW$ and $RCW$, i.e. $WRT = RW + RCW$, its pdf is the convolution of the pdf of $RW$ and the pdf of $RCW$:

$$f_{WRT}(t) = (f_{RW} \ast f_{RCW})(t) = \int_0^t f_{RW}(x)f_{RCW}(t-x)dx$$

The Laplace transform of $WRT$’s pdf is easier to develop since it is the product of Laplace transforms of $RW$’s pdf and $RCW$’s pdf. Thus applying the Laplace transform to Equations 1 and 2 we obtain:

$$F_{RW}(s) = \frac{(c^2_w - \alpha_w + 1)s + \alpha_w \eta_w}{(\alpha_w s + \eta_w)(s + \alpha_w \eta_w)}$$

$$F_{RCW}(s) = \frac{K_w}{K_w + s}$$

$$F_{WRT}(s) = F_{RW}(s)F_{RCW}(s)$$

$$= \frac{(c^2_w - \alpha_w + 1)s + \alpha_w \eta_w}{(\alpha_w s + \eta_w)(s + \alpha_w \eta_w)} \frac{K_w}{K_w + s}$$

Assuming $M_{WRT}(\cdot)$ is the moment generating function of $WRT$’s pdf, we can express $\eta_{WRT}$, i.e. the WLAN cell residence rate in the previous Equations, as the following:

$$\eta_{WRT} = \frac{1}{E(WRT)} = \frac{1}{M_{WRT}(0)} = \frac{1}{-\frac{dM_{WRT}(t)}{dt}|_{t=0}}$$

In fact, it is known that $M_{WRT}(t) = \frac{f_{WRT}(t)}{f_{WRT}(0)}$ for a continuous non-negative random variable $WRT$. Using
Equations 7 and 8, we formulate the mean CHT, i.e. the average WLAN occupancy time, of both types of calls:

\[
\frac{1}{\mu^\text{new}_{W}} = -(F^\text{new}_{WLAN})'(0) = \frac{\mu - \sigma_w \eta_w K_w}{\lambda_w \sigma_w + \sigma_h \gamma_w + \lambda_w \eta_w} (1 - F_{\text{WRT}}(\mu))
\]

\[
\frac{1}{\mu^\text{ho}_{W}} = -(F^\text{ho}_{WLAN})'(0) = \frac{1 - F_{\text{WRT}}(\mu)}{\mu}
\]

(10)

(11)

Similar Equations can be easily formulated for a 3G cell.

E. Expression of soft handoff probabilities

The soft handoff probability is defined as the probability that a call needs at least one more soft handoff during its remaining lifetime. We define the soft handoff as a call’s request of a new channel from the next AP/BS a period of time, called bi-casting period, before the call frees the occupied channel in the current AP/BS. Contrastingly, a hard handoff is defined as a call’s request of a new channel from the next AP/BS immediately after the call frees the occupied channel in the current AP/BS. In our model, the soft handoff is triggered after the elapsed residence time in the WLAN core region RW, by contrast to the hard handoff that is triggered later i.e. after the elapsed residence time in the whole region WRT=RW+RWC. Therefore, it is possible to reformulate the handoff probability expressions in Equation 12 and 16 in [11] to our problem. By replacing F_{\text{WRT}} by F_{\text{RW}} and \eta_{\text{WRT}} by \eta_{W}, we have the following expressions for soft handoff probabilities of newly initiated calls and handoff calls in the WLAN cell:

\[
P^\text{new}_{WLAN} = \frac{\eta_w}{\mu} (1 - F_{\text{RW}}(\mu))
\]

(12)

\[
P^\text{ho}_{WLAN} = F_{\text{RW}}(\mu)
\]

(13)

Note that the expressions developed for the CHT in Equation 11 and for the handoff probability in Equation 13 take into account the soft handoff region that we introduced in our model. However, the system performance in terms of blocking probabilities still have to be analyzed.

F. Estimating blocking probabilities

We model the voice capacity occupation in each WLAN (resp. 3G) cell using a bi-dimensional Markov chain illustrated in Figure 3 although the service time, i.e. the CHT, is not exponentially distributed since it is stated in [11] that the sum of hyper-exponential and Erlang distributions keeps the Markovian property. We can easily derive the Markov chain balance equations from Figure 3, then using the Gauss-Seidel approach on these equations we iteratively estimate state probabilities until the convergence to their respective steady-state probabilities. We use this approach since no closed-form expressions exist for these steady-state probabilities.

Following the admission policy described in Subsection II-C and using the estimated steady-state probabilities, we express the various blocking probabilities:

1) The new call blocking probability \( P^w_h = \Sigma_{j=0}^m P_h(i+j)P_i,j,k \).

2) The HHO failure probability : \( P^w_h = (1 - P_{\text{cov}}) \Sigma_{j=0}^m P_{i,j,w} + P_{\text{cov}} \) where \( P_{\text{cov}} \) is the WLAN out of coverage probability.

3) The VHO failure probability : \( P^w_{f,v} = P^c_f \) since upward VHO calls are admitted as newly initiated calls in the overlaying 3G cell due to the assumed loose coupling between 3G and WLAN.

4) The call dropping probability in the separate network (See Appendix): \( P^w_{d} = \frac{P_h f}{1 - P_h f (1 - P_f w)} \).

Similar equations can be formulated for the 3G cell. Considering the call continuity using VHO in the case of WLAN HHO failures, we develop more suitable expression for this dropping probability under the loosely coupled 3G/WLAN scheme:

\[
P^w_{d-c} = P^w_d P^c_f (1 + \frac{1 - P^c_w}{P^c_f})
\]

(14)

We provide our proof of this WLAN/3G dropping probability expression in the appendix.

G. Estimating handoff rates

In general, we use handoff rates as already given input parameters to the Gauss-Seidel algorithm that estimates the steady-state probabilities. However, in order to have more accuracy in our performance analysis, we estimate these handoff rates by applying the flow conservation fixed point equations that can be easily deduced from Figure 4. Starting with null handoff rates we iteratively evaluate these fixed-point equations and for each iteration we estimate the blocking probabilities using the Gauss-Seidel approach. We stop iterating on the fixed-point equations when all handoff rates converge to their respective steady values with a given resolution \( \varepsilon \). From Figure 4, we have the following fixed-point equations respectively for WLAN HHO rate \( v_w \), downward VHO rate
H. Voice Admission Control as an Optimization Problem

We know from Subsection II-F that the blocking probability function $\beta_{k}(.)$ is a WLAN-first strategy similar to the one analyzed in [7]. This strategy admits newly initiated calls and keeps handoff calls of multi-mode MS in the WLAN network before overflowing them to the overlaying 3G network. We noticed that for the common voice service most of the mobile operators try to distinguish from others by providing the lowest dropped calls. Therefore, the WLAN CAC has to minimize the new call blocking probability on WLAN while maintaining a voice call over the W-CDMA interface, in order to optimize resource utilization especially on the high-cost and highly loaded 3G cells. Since the soft handoff consumes duplicate resources to seamlessly handoff the call, in order to optimize resource utilization especially on the high-cost and highly loaded 3G cells, the soft handoff duration has to be minimized to its strict minimum. Standard implementations of soft handoff algorithms are simple and straightforward. However, a resource-efficient one is usually complex and CPU-consuming since it is challenging to compute accurate prediction estimates of the right time to trigger the soft handoff in a highly varying wireless mobile environment.

A. The Standard Static-Threshold Soft Handoff (STSHO)

While maintaining a voice call over the W-CDMA interface, the multi-mode MS monitors the received signal strength (RSS) on neighboring cells (3G and WLAN) and when it reaches a pre-defined RSS threshold (RSS$_{th}$ for the WLAN), it

$$v'_{w}, \text{ 3G HHO rate } v_{c}, \text{ and upward VHO rate } v'_{c}:$$

$$v_{w} = \lambda_{w}(1-P_{bc}^{w})h_{w}^{\text{new,wlan}} + v_{w}(1-P_{f}^{w})h_{w}^{\text{new,wlan}}$$

$$v_{c} = \lambda_{c}(1-P_{bc}^{c})h_{c}^{\text{new,3g}} + v_{c}(1-P_{c}^{w}+P_{c}^{w}h_{c}^{\text{new,wlan}})$$

$$v'_{w} = (1/m)P^{w}_{f}P^{w}_{c}$$

$$v'_{c} = mP^{c}_{f}P^{w}_{c}$$

III. The Soft Handoff Algorithm

Since the soft handoff consumes duplicate resources to seamlessly handoff the call, in order to optimize resource utilization especially on the high-cost and highly loaded 3G cells, the soft handoff duration has to be minimized to its strict minimum. Standard implementations of soft handoff algorithms are simple and straightforward. However, a resource-efficient one is usually complex and CPU-consuming since it is challenging to compute accurate prediction estimates of the right time to trigger the soft handoff in a highly varying wireless mobile environment.

A. The Standard Static-Threshold Soft Handoff (STSHO)

While maintaining a voice call over the W-CDMA interface, the multi-mode MS monitors the received signal strength (RSS) on neighboring cells (3G and WLAN) and when it reaches a pre-defined RSS threshold (RSS$_{th}$ for the WLAN), it
triggers the soft handoff to the next cell which will be preferably a WLAN cell if available. However, while maintaining the call over the IEEE 802.11 interface, it is more convenient to monitor the RSS on the current WLAN cell rather than neighboring WLAN cell, since switching channels to measure signals on neighboring WLAN cells during voice activity may degrade the voice QoS. Usually when the current RSS decreases below a certain level noted $RSS_{add}$, a soft handoff is triggered preferably to a neighboring WLAN cell, otherwise to an overlaying 3G cell. As in [18], after using an RSS averaging mechanism to smooth the shadowing fluctuations, the RSS level can be approximated by the latest interval during which the soft handoff (RESHO) algorithm for voice multi-mode MS is imminent to free the call on the next AP/BS before the RSS falls below $RSS_{out}$, which is the mean handoff signaling latency.

$$RSS(i\theta) = R_0 - 10n\log d(i\theta)\text{dBm}$$  \hspace{1cm} (20)
$$d^2(i\theta) = v^2(i\theta)\, t + d_0^2 - 2d_0\, v(i\theta)\cos(\alpha)$$  \hspace{1cm} (21)

We note $n$ the RSS slope index (typically 4 for micro-cells in the city). We assume the velocity model presented in Figure 6, where $v$ is the MS velocity intensity and $\alpha$ the angle of MS movement direction. We also assume time intervals with length $\theta$ short enough to have both $v$ and $\alpha$ constants during each instance $i$ of these time intervals.

B. The resource-efficient soft handoff (RESHO)

In order to minimize the use of highly expensive and highly loaded 3G cells, we propose using a resource-efficient soft handoff (RESHO) algorithm for voice multi-mode MS that estimates and adjusts the optimal RSS threshold of the VHO trigger depending on MS velocity and handoff signaling latency. We assume that the MS velocity and direction remain constant just before entering and inside the soft handoff region. This assumption is based on the fact that the time spent inside a RESHO region overcoming the handoff latency is short enough regarding the high WLAN residence time. Note that some more or less resource-efficient RESHO algorithms are presented in the literature [8], [9] but not evaluated as in our work. A possible design of a RESHO algorithm is the following. At the start of each time interval $i$, the RESHO algorithm estimates the RSS variation slope using the Equation 21 and the previously measured RSS values. Then it uses this estimation to predict the right time interval, noted $i_{add}$, for which a soft handoff have to be triggered. $i_{add}$ can be approximated by the latest interval during which the soft handoff have to be triggered in order to establish the end-to-end voice call on the next AP/BS before the RSS falls below the out of range $RSS_{out}$ which is imminent to free the call on the previous AP. This handoff latency of establishing the same end-to-end voice call on a different AP/BS can be signaled from a history table by the wireless access gateway, i.e. the equipment that loosely interconnects the two administrative domains of the respective core networks of 3G and WLAN. The detailed design of this algorithm is out of the scope of this paper.

C. Residence times for RESHO vs. STSHO

In order to quantify the performance gain achieved by using a RESHO mechanism as we described it, we express the mean residence times in the core WLAN region $t_{RW}$ and the whole WLAN region $t_{WRT}$ for the RESHO vs. STSHO algorithms. Recall that the STSHO is based on a pre-planned $RSS_{add} = \rho RSS_{out}$ configured for all MSs. We assume that each MS enters the WLAN cell at time $t = 0$ and leaves it at time $t = t_{WRT}$ with a constant velocity $v$ and a constant direction $\alpha$. We also assume that a hysteresis loop mechanism is implemented in order to prevent handoff ping-pong effects. Therefore we distinguish different RSS thresholds for moving in ($RSS_{in}$) and out of ($RSS_{out}$) the WLAN as illustrated in Figure 6.

$$RSS(0) = R_0 - 10n\log d_0 = RSS_{in}$$  \hspace{1cm} (22)
$$RSS(t_{RW}) = R_0 - 10n\log d(t_{RW}) = RSS_{add}$$  \hspace{1cm} (23)
$$RSS(t_{WRT}) = R_0 - 10n\log d(t_{WRT}) = RSS_{out}$$  \hspace{1cm} (24)

$$\eta(x) = \frac{v}{d_0 \cos \alpha + \sqrt{d_0^2 \cos^2 \alpha - d_0^2 + (e^{K_w \text{log}_e x})^2}}$$  \hspace{1cm} (25)
$$\eta_{WRT} = \frac{1}{t_{WRT}} = \eta(RSS_{out})$$  \hspace{1cm} (26)

Then, for STSHO and RESHO we have:

$$\eta_w = \frac{1}{t_{RW}} = \eta(RSS_{add})$$  \hspace{1cm} (27)
$$\sigma_w = \frac{1}{t_{sig}} = \frac{1}{t_{WRT} - (r/K_w) - t_{RW}}$$  \hspace{1cm} (28)
$$\eta_w^* = \frac{1}{t_{WRT} - (r/K_w) - t_{sig}}$$  \hspace{1cm} (29)
$$\sigma_w^* = \frac{1}{t_{sig}}$$  \hspace{1cm} (30)

For different velocities and a given configured RSS threshold $RSS_{add}$, it is possible to use the previous equations to compute $\eta_w$ and $\sigma_w$ for the STSHO algorithm and $\eta_w^*$ for the RESHO algorithm. $1/\sigma_w^*$ is the mean handoff signaling latency which is assumed independent from the MS velocity.

IV. PERFORMANCE RESULTS

In this Section we present the performance results provided by the proposed models and algorithms. Table II presents the values used for the input parameters.
### A. Analytical model validation

In order to validate our blocking probability analysis, we have developed our own discrete event simulator using C++. We assumed a 3G ubiquitous coverage area divided into several hexagonal cells, each cell has six neighbors’ cells. In order to cover as many cells as possible and eliminate the edge effect not present in a ubiquitous coverage, we used the well known wrap around model which includes in our case 49 cells. By wrapping around the 49 cells, the handoff traffic which flows out of the coverage will flow into the area again. This prevents from losing calls. We assumed a WLAN area with limited coverage overlaying each 3G cell. Exactly $m = 7$ WLAN cells overlay one single 3G cell with a planned out of coverage probability $P_{\text{cov}}$. In each cell (either 3G or WLAN), new call arrivals are generated according to the Poisson distribution with their respective mean rates. The handoff call arrivals are triggered in the simulator, rather than generated, when their respective generated CHT durations elapse. Our presented simulation results are statistical values that are averaged over the required number of runs to achieve a 95% confidence level of less than 0.001 of confidence interval width. The purpose of the simulation is twofold. First, since no appropriate analytical model characterizing the handoff traffic exists in the literature and therefore simulation is the only way to estimate it accurately. Second, even though simulations consume excessive CPU cycles, it is a powerful way to check if all the blocking and dropping probabilities estimated using the analytical model are still valid in the real simulated system.

From Figure 7 we notice that the 3G cell new call blocking and ongoing call dropping probabilities estimated using our proposed analytical model are almost equal to the ones estimated using simulation. Similarly, from Figure 7 we observe that the analysis results agrees with the simulation results for the WLAN cell. Furthermore, from these Figures we also observe that the minB algorithm adjusts the optimal 3G cell LFGC $g_{\text{c}}$ (respectively the optimal WLAN cell LFGC $g_{\text{w}}$) depending on the new call arrival rate such that it minimizes the new call blocking rate while maintaining a hard QoS constraint on a call dropping rate not exceeding $10^{-2}$. We notice that due to the lower handoff traffic intensity in WLAN cells the LFGC starts to increase only at a higher new call arrival rate of $2.75$ calls/mn for WLAN vs. $1.75$ calls/mn for 3G.

### B. Resource-efficient soft handoff performance

From results presented in Figures 8, 10 and 9 we obviously notice that under all mobility schemes in the WLAN area (MS velocity with different mean from 0.5 m/s to 4 m/s and different variability from 1 to 6) the RESHO algorithm outperforms the STSHO algorithm in terms of optimal CAC performance metrics. This is due to the fact that in average, duplicate capacity resources for the soft handoff are occupied for shorter time when using the RESHO algorithm. Note that the call dropping probability of calls initiated in the 3G network is not presented on the results since it is constantly equal to the QoS threshold ($P_{d,\text{max}}^c = 10^{-2}$) that is due to the operation of the 3G network in a high load region. Surprisingly in Figure 10 increasing the mean velocity in the WLAN area increases the blocking probability of new calls on WLAN cells which seems counter-intuitive since lower channel occupation duration leaves more available capacity for new calls. However, this fact can be explained by the increased WLAN horizontal soft handoff traffic which is re-injected in the WLAN cells with more duplicate resources occupied and thus increasing the WLAN load more than a higher velocity decreases it. Furthermore, we notice that increasing the velocity variability $\sigma_v$ in the WLAN area decreases the new call blocking probability on WLAN network. In fact, since a larger velocity variability $\sigma_v$ indicates that more users staying within the WLAN for shorter time (fast users) and less users staying within the WLAN for longer time (slow users), a large fraction of voice calls initiated in the WLAN area can be carried by the less occupied WLAN which consequently relieves the 3G cell from the overflowing traffic.

More interestingly, it is observed that the RESHO algorithm provides much higher performance gain than the standard STSHO algorithm for lower MS velocities. In fact, as noticed in Figures 8 and 9 the reduction ratios for both 3G new call blocking probability and WLAN initiated call dropping probability are significantly increased when the MS velocity is low. This is due to the fact that when the MS velocity decreases, the deviation of the residence time in the static STSHO soft handoff region from the residence time in the

---

### TABLE II

**SYSTEM INPUT PARAMETERS**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>3.3</td>
<td>$P_{d,\text{max}}^c$</td>
<td>1%</td>
</tr>
<tr>
<td>RSS$_{\text{net}}$</td>
<td>-85 dBm</td>
<td>$c$</td>
<td>13</td>
</tr>
<tr>
<td>RSS$_{\text{in}}$</td>
<td>-80 dBm</td>
<td>$c_w$</td>
<td>15</td>
</tr>
<tr>
<td>RSS$_{\text{adj}}$</td>
<td>-60 dBm</td>
<td>$(\mu_i)^{-1}$</td>
<td>4 mn</td>
</tr>
<tr>
<td>$R_0$</td>
<td>70.28 dBm</td>
<td>$K_c$ ($K_w$)</td>
<td>60 ms</td>
</tr>
<tr>
<td>$d_h$</td>
<td>95 m</td>
<td>$pe_i$ ($pc_i$)</td>
<td>5%</td>
</tr>
<tr>
<td>$K_w$</td>
<td>60 msec</td>
<td>$(\eta_i)^{-1}$</td>
<td>2 mn</td>
</tr>
<tr>
<td>$m$</td>
<td>7</td>
<td>$(\sigma_v)^{-1}$</td>
<td>600 ms</td>
</tr>
<tr>
<td>$P_{\text{cov}}$</td>
<td>20%</td>
<td>$\varepsilon$</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

---

![Fig. 7. Validation of the proposed analytical model for 3G and WLAN](image-url)
dynamic RESHO soft handoff region increases. Besides, when duplicate resources are consumed for less time the blocking probabilities are highly reduced.

In addition, besides a low MS velocity inducing a better performance gain, when the velocity variability is high ($\sigma_v = 6$) the performance gain for the 3G new call blocking probability is even higher. This is due to the fact that the increase of velocity variability induces more temporal statistical multiplexing in the WLAN capacity occupation inducing less WLAN blocking and less WLAN handoff failures and consequently less overflowing upward VHO calls traffic and less 3G new call blocking probability.

Since a low-mean high-variance velocity profile typically characterizes the WLAN mobility environment [10], it may push WLAN designers to implement the RESHO algorithm if its expected performance gain estimated by our proposed model is appreciated. Furthermore, the significant performance gain provided by the RESHO algorithm in particularly low mobility environments make the algorithm a somewhat less difficult to implement. In fact, when the MS mean velocity remains low, more samples can be used to predict a more accurate RESHO residence time.

V. CONCLUSION

In this paper we presented a novel analytical model in order to evaluate the new call blocking probability and the call dropping probability of the optimal voice admission control algorithm redefined in loosely coupled 3G/WLAN networks. This model uses more accurate residence time distributions and takes into account the soft handoff region for each 3G/WLAN cell for which duplicate resources are allocated. Results show that the delimitation of the soft handoff region using a more or less efficient soft handoff algorithm has a big impact on the CAC performance metrics. Particularly, it is observed that the new call blocking probability reduction gained by using a resource-efficient soft handoff (RESHO) algorithm compared to a static-threshold soft handoff (STSHO) algorithm is largely increased when MS velocities have low mean and high variability; which typically characterizes the WLAN environment. Our proposed model, validated by simulations, can help network designers to figure out if the gained performance from a challenging RESHO algorithm in different mobility environments is worth to implement it.

APPENDIX

Derivation of the 3G/WLAN dropping probabilities

Assuming that call dropping can only be avoided by successful obligatory handoffs, i.e. HHO and UVHO, the expression of dropping probability of a call initiated in the 3G, noted $P_{d}^c$, is already known in cellular networks:

$$ P_{d}^c = \sum_{i=1}^{\infty} (P_h^i)^{\alpha} (1 - P_f^i)^{\alpha - 1} P_f^z = \frac{P_h^z P_f^c}{1 - P_f^z} $$

(31)

Where $P_h^i$ is the 3G handoff probability and $P_f^z$ is the 3G handoff failure probability. However, the expression of dropping probability of a call initiated in the overlaying WLAN, noted $P_{d}^w$, needs to be developed. Now let $P_e = 1 - P_{d}^c$ the probability that a voice call that is not initially blocked
by a WLAN will be normally completed without any handoff failure. $P_c$ is the sum of the probabilities of the following events:

- The call is completed normally in the first WLAN cell;
- The call is completed normally in a WLAN cell after one or more WLAN-to-WLAN HHs which are all successful;
- The call is completed normally in a 3G cell after one or more WLAN-to-3G UVHs, or eventual subsequent 3G-to-3G HHs which are all successful.

We assume that an ongoing call moving to the WLAN from the overlaying 3G cell (with a small probability $P_{w-c}$ as an optional continuity to the previous event) can not be dropped since the overlaying 3G cell won't free its reserved channel before ensuring the successful soft DVHO to the WLAN where it is assumed to be completed normally. Thus:

$$P_c = (1 - P_{hw}) + \sum_{i=1}^{\infty} \left( \sum_{j=1}^{\infty} (P_{h}^w)^{j-1} P_{c} (1 - P_{h}^w) \right)$$

$$P_{cw} = \frac{1}{1 - P_{c}}$$

$$P_{cw}^w = \frac{1}{1 - P_{h}^w}$$

$$P_{cw}^c = \frac{1}{1 - P_{c}}$$

$$P_{d-w} = \frac{P_{h}^w P_{f}^w (P_{h}^w P_{f}^w + P_{c} - P_{h}^w P_{f}^c)}{(1 - P_{h}^w + P_{h}^w P_{f}^w) (1 - P_{c} + P_{h}^w P_{f}^c)}$$

$$P_{d-c} = \frac{P_{h}^c P_{f}^c (1 + \frac{1}{P_{h}^c} - \frac{P_{c}}{P_{f}^c})}{P_{f}^c}$$

If WLAN-to-3G UVHO calls are admitted as 3G-to-3G HH calls which can be done in a tightly coupled 3G/WLAN, i.e. by replacing $P_{h}$ with $P_{f}$ in the last Equation, then $P_{d-w}^w = P_{d}^w P_{h}^c P_{f}^c$. Recall that $P_{d}^w$ and $P_{d}^c$ are respectively the dropping probabilities in the separate WLAN and 3G networks respectively.

REFERENCES


Racha Ben Ali (S’02) received the M.Sc. degree (2003) and the Ph.D. degree (2008) in Computer Engineering from Ecole Polytechnique de Montreal, Qu’bec, Canada. He was a computer network architect and engineer at Tele-University of Quebec from 2006 to 2007. Previous to that, he worked as a Graduate Research Assistant with the Mobile Computing and Networking Research Laboratory (LARIM) and with Ericsson Research Canada. The current focus of his research interests is on radio resource management, call admission control, heterogeneous mobile networks interworking, performance evaluation and quality of service (QoS) provisioning for next-generation networks.

Sanueil Pierre (SM’97) received the B.Eng. degree in Civil Engineering in 1981 from Ecole Polytechnique de Montreal, Qu’bec, Canada, the B.Sc. and M.Sc. degrees in Mathematics and Computer Science in 1984 and 1985, respectively, from the UQAM, Montreal, the M.Sc. degree in Economics in 1987 from the University of Montreal, and the Ph.D. degree in Electrical Engineering in 1991 from Ecole Polytechnique de Montreal. From 1987 to 1998, he was a Professor at the University of Qu’bec at Trois-Rivi’eres prior to joining the T’el’e-Universit’e of Qu’bec, an Adjunct Professor at Universit’e Laval, Ste-Foy, Qu’bec, an Invited Professor at the Swiss Federal Institute of Technology, Lausanne, Switzerland, and the the Universit’e Paris 7, France.

He is currently a Professor of Computer Engineering at Ecole Polytechnique de Montreal, where he is Director of the Mobile Computing and Networking Research Laboratory (LARIM), Chairholder of the NSERC/Ericsson Chair in Next Generations and Mobile Networking Systems, and the Director of the Mobile Computing and Networking Research Group (GIM). His research interests include radio resource management, call admission control, heterogeneous mobile networks interworking, performance evaluation and quality of service (QoS) provisioning for next-generation networks.