Performance Modeling for Heterogeneous Wireless Networks with Multiservice Overflow Traffic

Qian Huang, King-Tim Ko and Villy Bæk Iversen

Abstract—Performance modeling is important for the purpose of developing efficient dimensioning tools for large complicated networks. But it is difficult to achieve in heterogeneous wireless networks, where different networks have different statistical characteristics in service and traffic models. Multiservice loss analysis based on multi-dimensional Markov chain becomes intractable in these networks due to intensive computations required. This paper focuses on performance modeling for heterogeneous wireless networks based on a hierarchical overlay infrastructure. A method based on decomposition of the correlated traffic is used to achieve an approximate performance modeling for multiservice in hierarchical heterogeneous wireless networks with overflow traffic. The accuracy of the approximate performance obtained by our proposed modeling is verified by simulations.

Index Terms—Hierarchical overlay networks, femtocells, multiservice loss model, overflow traffic.

I. INTRODUCTION

Future broadband wireless networks aim to provide all-IP based access through integrating various existing wireless access technologies, e.g. WiFi, WiMAX, 3G and LTE. To facilitate internetworking between these heterogeneous networks, hierarchical overlay infrastructure has been considered as a possible solution, e.g. overlaying GPRS/WiMAX on top of WiFi access points [1], or cellular systems over femtocells [2]. When mobile users enter the uncovered areas of cellular base stations, on-going traffic can overflow to neighboring open-access femtocell networks for continuous connection; or when call requests are rejected by femtocells due to capacity limit, the blocked calls can overflow to the covering GPRS/WiMAX.

Efficient and accurate modeling is crucial for performance evaluation and network dimensioning in large networks. However, it is a difficult issue in hierarchical heterogeneous wireless networks because of the following aspects.

Based on the hierarchical overlay infrastructure, the traffic unable to be accommodated in a lower tier network due to limited capacity is allowed to overflow to a higher tier network for possible service [3]. This scheme improves the capacity utilization of overlay networks, and also reduces the blocking probability since the available capacity of higher tier networks can be shared amongst the blocked traffic in lower tier networks. The problem is that the inter-tier overflow traffic is bursty in nature [4], and traditional analysis for Poisson traffic in multiservice loss systems is no longer available for the non-Poisson overflow traffic. Existing approximation methods include the use of Markov-modulated Poisson process (MMPP) or its variants to approximate the overflow process [5], [6], [7]. Then, approximate loss performance of overflow traffic can be obtained by solving the Markov chain models. The major concern with the MMPP approximation is its still high complexity involved in modeling the multiservice overflow traffic.

Overflow traffic is generally characterized by its first two moments [4], i.e. mean and variance of the offered overflow traffic intensity. The mean of a specific class of overflow traffic can be directly obtained as the equivalent blocked traffic of this class in its overflowed network group. The ratio of variance to mean is defined as peakedness, which indicates the bursty nature of the overflow traffic. In homogeneous hierarchical networks where all network groups at different tiers have identical statistical characteristics, e.g. the two-tier overlay system discussed in [7], the groups in each tier are statistically identical; a specific class of overflow traffic from a low tier to a higher tier holds identical statistical moments in these two tiers. Unlike homogeneous networks, heterogeneous networks give rise to different statistical characteristics of input traffic, mobility model and service time distribution for different networks at different tiers. The statistical moments of the overflow traffic from one tier network overflow to another tier are required to be redefined. Hence, another key problem is to model overflow traffic in heterogeneous networks. This problem is also involved in the speed-sensitive handoff and traffic overflow between cellular systems and femtocells, as one of the major technical challenges [8].

This paper aims to solve multiservice performance modeling for hierarchical heterogeneous wireless networks. We consider the statistical heterogeneity in service, traffic and mobility models at different tiers, when determining the statistical moments of the inter-tier overflow traffic. We calculate the loss performance at each tier offered with the superposition of input Poisson and overflow non-Poisson traffic, using an approximate analysis based on decomposition of the correlated traffic. Our system model is described in Section II. The related work is reviewed in Section III. The proposed approximate performance modeling for the hierarchical heterogeneous networks is presented in Section IV, followed by performance evaluation of the proposed approximate model in a multi-tier heterogeneous system and verification by simulations in Section V. The conclusions are presented in Section VI.
II. SYSTEM MODEL

Consider a hierarchical heterogeneous system with $L \geq 2$ tiers supporting multiservice and mobility based on an overlay infrastructure. The networks may have different transmission technologies, capacity limits and therefore different statistical characteristics for a specific traffic class, e.g. service time distribution and mobility model, in different tiers. The networks in the same tier are assumed to be homogeneous and have identical statistical characteristics. It is reasonable to assume that the networks at the higher tiers have larger coverage and capacity than the networks at the lower tiers. Blocked traffic is allowed to overflow between heterogeneous networks at different tiers to improve capacity utilization and blocking probability. The traffic input to any tier network comprises new calls from local users, handoff calls from handoff users of the neighboring network groups in the same tier, as well as overflow calls from users in other tiers.

New calls blocked in a tier $l$ network due to capacity limit can be overflowed to the networks with available capacity at the higher tiers. A speed-sensitive overflow control scheme [1], [9] is used to tackle the handoff calls. Mobiles in the networks are classified into fast-speed and slow-speed in terms of their mean velocities. The blocked handoff calls of fast-speed users in a tier $l$ network are overflowed to the high tier $l+1$ network with larger coverage and capacity (e.g. from femtocells to WiMAX). If the overflowed fast-speed calls are rejected at tier $l+1$, further overflow to high tier $l+2$ is attempted. The identical strategy can be used to tackle slow-speed users’ handoff, but the blocked calls are overflowed to the lower tier networks (e.g. from GPRS/WiMAX to WiFi/femtocells) if there is available capacity. Call (new or handoff call) blocking only occurs when there is no available capacity in any appropriate tier, hence both call blocking and dropping probabilities are reduced. Comparing with slow-speed users, fast-speed users have higher number of handoffs in smaller networks (e.g. femtocells), therefore higher possibility of handoff failure. The speed-sensitive scheme reduces the signaling cost and handoff failure since frequent handoff is avoided.

The traffic involved in our performance modeling is defined as follows. In the following notations, subscript $i = 0$ is used as the denotation related to new call traffic and subscript $i = 1$ as the denotation related to handoff traffic. For class $k$ ($1 \leq k \leq K$) service in tier $l$ ($l \geq 1$) network group, we have

1) local new call traffic with offered load intensity $\rho_0^{(k,l)}$;
2) local handoff traffic with offered load intensity $\omega_0^{(k,l)}$ due to the handoff of the admitted calls within tier $l$;
3) overflowed new call traffic with offered load intensity $\rho_0^{(k,l-1)}$ contributed by fast-speed users blocked in tier $l-1$ groups and $\omega_0^{(k,l+1)}$ by slow-speed users blocked in tier $l+1$ groups;
4) overflowed handoff call traffic with offered load intensity $\rho_1^{(k,l-1)}$ contributed by fast-speed users’ handoffs blocked in tier $l-1$ groups and $\omega_1^{(k,l+1)}$ by slow-speed users’ handoffs blocked in tier $l+1$ groups.

It is assumed that each tier $l$ network group overlays $J$ tier $l-1$ network groups. The mean call arrival rate of class $k$ traffic to a tier $l$ group, denoted as $\Lambda_i^{(k,l)}$, is determined by

$$
\Lambda_i^{(k,l)} = \alpha_i^{(k,l)} + J \rho_1^{(k,l-1)} + \omega_i^{(k,l+1)}, \quad \text{for} \; i = 0, 1. \; (1)
$$

In Fig. 1, $Z_i^{(k,l)}$ is the peakness of class $k$ traffic to tier $l$ network, $d_k$ is the bandwidth required by each class $k$ call.

New call traffic is assumed to be Poisson. Handoff traffic within the same tier network can also be modeled as a Poisson process [10]. The handoff traffic intensity offered to a network can be determined by a statistical equilibrium [10], that is the average departing handoff traffic out of a cell is equal to the average arriving handoff traffic into the cell, based on the assumption that all neighboring cells are homogeneous and identical. By this assumption, the intensity of type $k$ handoff traffic offered to a tier $l$ network group, $\alpha_i^{(k,l)}$, is derived as

$$
\alpha_i^{(k,l)} = \rho_0^{(k,l)} + \rho_0^{(k,l-1)} + \omega_0^{(k,l+1)}(1 - B_0^{(k,l)}), \quad \text{for} \; i = 0, \quad (2)
$$

Approximate loss modeling for multi-rate non-Poisson traffic: Consider a trunk group denoted as $(N; A, Z, d)$ with $N$ servers, offered by a non-Poisson source traffic with intensity $A$ and peakness $Z$, and a unit service time for each admitted call which occupies $d$ servers. It has been demonstrated that the approximate loss performance of this multi-rate non-Poisson traffic in the $N$-server trunk group can be obtained in an equivalent $N$-server trunk group offered by a Poisson

Fig. 1. Hierarchical overflow system model.
traffic with intensity $A/Z$ and each admitted call occupying $dZ$ servers [7]:

$$B(N; A, Z, d) \equiv E(N; \frac{A}{Z}, 1, dZ),$$

(3)

where $B(\ )$ denotes call blocking probability, $E(\ )$ denotes the Erlang loss function. We refer to the approximation (3) as multi-rate Hayward’s approximation; it can be regarded as the extension of the traditional Hayward’s approximation for single-rate non-Poisson traffic [11].

**Semi-equivalent loss systems:** The principle of Hayward’s Approximation is based on mapping a original trunk group $(N; A, Z)$ into an equivalent model comprising $Z$ independent identical subgroups in parallel [11]. Each subgroup has $N/Z$ servers. The calls of the non-Poisson source traffic $(A, Z)$ are evenly scheduled into each subgroup so that the traffic offered to each subgroup is ensured to be Poisson with intensity $A/Z$. The same service time is assumed for a call in the original trunk group and in the subgroups. It is verified that the resulted subgroup, denoted as $(N/Z; A/Z, 1)$, has equivalent call blocking probability to the trunk group $(N; A, Z)$ [11]. This derivation was used as Hayward’s approximation. We here show that the two system models $(N; A, Z)$ and $(N/Z; A/Z, 1)$ have different statistical characteristics of their overflow traffic. We refer to such two systems as a pair of **semi-equivalent loss systems**. The proof is presented in the following.

The actual values of the moments of overflow traffic can be obtained in a “fictitious” overflow group with infinite number of servers in the secondary trunk group to admit the overflow calls from the primary trunk group [4]. Let $p_x(i, j)$ denote the probability of subgroup $x$ at the state that there are $i$ new calls in subgroup $x$ and $j$ overflow calls in its infinite-server overflow group, $0 \leq i \leq N/Z$ and $j \geq 0$. Since all subgroups are independent of each and identical, for any two subgroups $x$ and $x’ (x, x’ = 1, 2, ..., Z, x’ \neq x)$ we have

$$p_x(i, j) = p_x(i, j).$$

(4)

Let $m_{sub}$, $v_{sub}$ and $z_{sub}$ denote mean, variance and peakedness of the overflow traffic from the trunk group $(N/Z; A/Z, 1)$, i.e. a subgroup $x (x = 1, 2, ..., Z)$. They are obtained as

$$m_{sub} = \sum_{i=0}^{N/Z} \sum_{j=0}^{\infty} j \cdot p_x(i, j),$$

(5)

$$v_{sub} = \sum_{i=0}^{N/Z} \sum_{j=0}^{\infty} j^2 \cdot p_x(i, j) - m_{sub}^2,$$

(6)

$$z_{sub} = v_{sub}/m_{sub},$$

(7)

The overflow traffic out of the original trunk group $(N; A, Z)$ is equivalent to the superposition of the overflow traffic from all $Z$ subgroups. Let $p(k, l)$ denote the probability of the original trunk group at the state that there are $k$ new calls in the trunk group $(N; A, Z)$ and $l$ overflow calls in the infinite-server overflow group, $0 \leq k \leq N$, $l \geq 0$. Since the calls of the non-Poisson source traffic are evenly scheduled to each subgroup, we have

$$p_x(i, j) = p(Z \cdot i, Z \cdot j) = p(k, l)$$

(8)

for $k = Z \cdot i, l = Z \cdot j$, and $1 \leq x \leq Z$.

Based on (5), (6) and (8), the mean and variance of the overflow traffic from the trunk group $(N; A, Z)$, denoted as $M_{ov}$ and $V_{ov}$, are derived as

$$M_{ov} = \sum_{k=0}^{N} \sum_{l=0}^{\infty} l \cdot p(k, l)$$

$$= \sum_{k=0}^{N} \sum_{l=0}^{\infty} Z \cdot j \cdot p(Z \cdot i, Z \cdot j)$$

$$= Z \sum_{i=0}^{N/Z} \sum_{j=0}^{\infty} j \cdot p_x(i, j) = Z \cdot m_{sub},$$

(9)

$$V_{ov} = \sum_{k=0}^{N} \sum_{l=0}^{\infty} l^2 \cdot p(k, l) - M_{ov}^2$$

$$= \sum_{k=0}^{N/Z} \sum_{l=0}^{\infty} Z^2 \cdot j^2 \cdot p_x(i, j) - Z^2 m_{sub}^2$$

$$= Z^2 \cdot v_{sub},$$

(10)

and the peakedness of the overflow traffic from the trunk group $(N; A, Z)$, denoted as $Z_{ov}$, is given by

$$Z_{ov} = V_{ov}/M_{ov} = Z \cdot z_{sub}.$$

(11)

**IV. HIERARCHICAL HETEROGENEOUS PERFORMANCE MODELING**

For simplicity, we here elaborate our proposed performance modeling in the case of upward overflow traffic from lower tier to higher tier networks. The same modeling method is also available for downward overflow traffic.

**A. Traffic input to tier $l$ network group**

Referring to Fig. 1, where $(A^{(k,l)}, Z^{(k,l)}, d_k)$ represents the input traffic to tier $l$ network. It is characterized by the offered load intensity $A^{(k,l)}$, the peakedness $Z^{(k,l)}$ ($i = 0, 1$) and the required bandwidth $d_k$ for each class $k$ call. Here, $i = 0$ is used to represent new call traffic and $i = 1$ for handoff call traffic. In particular, Poisson traffic has peakedness equal to one. We here consider a general performance modeling; the peakedness of the input traffic can be larger than one, since the overflow traffic has peakedness $> 1$.

Our previous investigation has demonstrated the feasibility of tackling the non-Poisson traffic by mapping it to an equivalent system offered with the approximated Poisson traffic obtained by the multi-rate Hayward’s approximation (3) [7]. This approach was verified to achieve sufficiently accurate performance evaluation for homogeneous overlay networks, with
lower complexity compared with the methods based on multi-
dimensional Markov chain models [3], [12]. Using the same
approach, the loss performance of the traffic \( (\Lambda_i^{(k,l)}, Z_i^{(k,l)}, d_k) \) 
\((i = 0, 1)\) for class \( k, 1 \leq k \leq K_i \), in tier \( l \) network can be
approximately obtained in an equivalent system offered with
Poisson traffic \( \left( \frac{\Lambda_i^{(k,l)}}{Z_i^{(k,l)}}, 1, d_k Z_i^{(k,l)} \right) \), given that this equivalent
system has the same statistical characteristics of user mobility
model and service time distribution as in the original tier
\( l \) network for class \( k \) service. The loss performance of the
equivalent system can be obtained by the traditional methods
for Poisson traffic loss analysis.

B. Peakedness of overflow traffic in heterogeneous scenarios

For the considered upward overflow traffic from tier \( l - 1 \)
network to tier \( l \) network in Fig. 1, its mean is equivalent
to the blocked traffic load in tier \( l - 1 \) network; it can be
obtained from the blocking probability in the tier \( l - 1 \) group.
Let \( B_0^{(k,l-1)} \) and \( B_1^{(k,l-1)} \) denote the obtained blocking prob-
abilities for class \( k \) new calls and handoff calls respectively,
in the tier \( l - 1 \) network. The intensities of the blocked traffic
load for class \( k \) new calls and handoff calls are obtained as
\( \Lambda_0^{(k,l-1)} B_0^{(k,l-1)} \) and \( \Lambda_1^{(k,l-1)} B_1^{(k,l-1)} \) respectively.

The actual peakedness of the overflow traffic can be ob-
tained from the statistical distribution of the number of overflow
calls in an overflow group assumed to have infinite number of servers for the overflow calls. However, it is imprac-
ticable to solve a Markov chain model of infinite number of
states. A reasonable solution is to calculate the approximate
statistical moments of overflow traffic in a truncated overflow
group [9], [12], but the computations involved in large multi-
dimensional Markov chain for multiservice overflow are still
very extensive. An earlier investigation [13] has demonstrated
that given a single Poisson traffic with intensity \( a \) input to a
trunk group with \( N \) servers, the variance of the overflow
traffic from the trunk group can be obtained by

\[
v \approx m \left( 1 - m + \frac{a}{N + 1 - a + m} \right),
\]

(12)

where \( m \) denotes the mean of the overflow traffic. This
derivation gives rise to an alternative method to determine
the peakedness of the overflow traffic in a computationally
efficient way.

To obtain the variance of each individual overflow traffic
from different service classes from a tier \( l - 1 \) network, we
introduce the decomposition method used in [7]. A set of
independent trunk groups is assumed to accommodate the
individual traffic flow of new and handoff calls to the tier
\( l - 1 \) network, by assuming

1) each trunk group admits only one input traffic flow of
the equivalent tier \( l - 1 \) network obtained by the multi-
rate Hayward’s approximation;
2) the calls of a specific service class are given the identical
service time in the independent trunk group as they are in
tier \( l - 1 \) network;
3) the link capacity of the independent trunk group for a
specific traffic is determined under the constraint that
the traffic in the independent trunk group has identical
blocking probability as in tier \( l - 1 \) network.

Let \( \beta_i^{(k,l-1)} \) denote the capacity of the independent trunk
group for class \( k \) calls in tier \( l - 1 \) network. The independent
trunk group for tier \( l - 1 \) network group is denoted as
\( \left( \beta_i^{(k,l-1)}, \frac{\Lambda_i^{(k,l-1)}}{Z_i^{(k,l-1)}}, 1, d_k Z_i^{(k,l-1)} \right) \). Let \( \hat{m}_i^{(k,l-1)} \) and \( \hat{z}_i^{(k,l-1)} \) denote mean and peakedness of the
overflow traffic from the independent trunk group. Here, \( i = 0 \)
is for denoting new calls and \( i = 1 \) for handoff calls. The mean \( \hat{m}_i^{(k,l-1)} \) is equivalent
to the blocked traffic load in the independent trunk group, i.e.

\[
\hat{m}_i^{(k,l-1)} = B_i^{(k,l-1)} \cdot \frac{\Lambda_i^{(k,l-1)}}{Z_i^{(k,l-1)}}.
\]

(13)

The peakedness \( \hat{z}_i^{(k,l-1)} \) can be obtained by (12) as follows:

\[
\hat{z}_i^{(k,l-1)} \approx 1 - \hat{m}_i^{(k,l-1)} + \frac{\Lambda_i^{(k,l-1)}}{Z_i^{(k,l-1)} + 1} - \frac{\Lambda_i^{(k,l-1)}}{Z_i^{(k,l-1)}} + \hat{m}_i^{(k,l-1)}.
\]

(14)

Equation (14) gives the approximate peakedness of the
overflow traffic from the equivalent system for tier \( l - 1 \)
network obtained by the multi-rate Hayward’s approximation.
Regarding to the aforementioned constraint 3) of the
decomposition method, \( \beta_i^{(k,l-1)} \) is determined by

\[
E(\beta_i^{(k,l-1)}, \frac{\Lambda_i^{(k,l-1)}}{Z_i^{(k,l-1)}}, 1, d_k Z_i^{(k,l-1)}) = B_i^{(k,l-1)}
\]

(15)

The values of \( \beta_i^{(k,l-1)} \) for \( i = 0, 1 \) can be calculated by the
Erlang-B formula in an iterative fashion:

\[
E(j, a) = \frac{a E(j - 1, a)}{j + a E(j - 1, a)} \forall j = 1, 2, \ldots, n
\]

(16)

where \( a \) is the offered traffic load in Erlang, \( n \) is the number of
servers, \( E(0, a) = 1 \).

The mean and peakedness obtained in (13) and (14) are for
the overflow traffic from tier \( l - 1 \) network, related to the mean
call service times \( 1/\mu_i^{(k,l-1)} \) (\( i = 0 \)) for new and handoff
calls in tier \( l - 1 \) network. Once the overflow calls from tier
\( l - 1 \) network, they are subject to another service time distribution with mean \( 1/\mu_i^{(k,l)} \) in tier
\( l \) network, due to different capacity in different tier networks.
Such heterogeneity can also be observed in a homogeneous
overlay mobile network, e.g., the micro/macro mobile cellular
overlay networks considered in [3] and [12]. In these networks,
since the cells in different tiers have different cell size, a
specific mobile call can have different cell dwelling time
distributions in different tiers. This gives different service time
distributions for this mobile call in different tiers. However,
this statistical heterogeneity on overflow traffic has not been
taken into account in other current investigations.

For simplicity but not compromising the heterogeneity in
our considered performance modeling, we assume that for
a specific service class of mobile calls, their call service times in different tiers follow an identical distribution but with different mean call service times. In such circumstances, for the overflow traffic from tier $l-1$ network to tier $l$ network, its mean and peakedness input to tier $l$ network can be obtained according to the service time distributions in tier $l$ network by a recursive method. Define $\epsilon_i^{(k,l)}$ as the ratio of mean call service time in tier $l$ group to that in tier $l-1$ group for class $k$ calls, i.e., $\epsilon_i^{(k,l)} = \mu_i^{(k,l-1)}/\mu_i^{(k,l-1)}$, $i = 0, 1$.

Let $\tilde{m}_i^{(k,l)} (i = 0, 1)$ be the mean of the upward overflow traffic from tier $l-1$ to tier $l$ network. It is obtained as

$$\tilde{m}_i^{(k,l)} = \Lambda_i^{(k,l-1)}/Z_i^{(k,l-1)} \epsilon_i^{(k,l)}.$$  

(17)

Let $\tilde{z}_i^{(k,l)} (i = 0, 1)$ be the peakedness of the upward overflow traffic from tier $l-1$ to tier $l$ network. By the similar iterative algorithm of [14], $\tilde{z}_i^{(k,l)}$ can be derived as

$$\tilde{z}_i^{(k,l)} \approx 1 - \frac{\Lambda_i^{(k,l-1)} B_i^{(k,l-1)}}{Z_i^{(k,l-1)} \epsilon_i^{(k,l)}} + \frac{\Lambda_i^{(k,l-1)} g_i(\beta_i^{(k,l-1)})}{Z_i^{(k,l-1)} \epsilon_i^{(k,l)} g_i},$$  

(18)

where $g_i(\beta_i^{(k,l-1)})$ is defined by the following recursion, with the initial definitions $g_i(-1) = 0$ and $g_i(0) = 1$ for $i = 0, 1$:

$$g_i(x) = \frac{x}{\Lambda_i^{(k,l-1)} Z_i^{(k,l-1)}} + x - 1(\epsilon_i^{(k,l)}) g_i(x-1)$$  

$$- \Lambda_i^{(k,l-1)} Z_i^{(k,l-1)} x g_i(x-2), 1 \leq x \leq \beta_i^{(k,l-1)},$$  

(19)

and $\hat{g}_i$ is given by $\hat{g}_i = \sum_{x=0}^{\beta_i^{(k,l-1)}/d_k Z_i^{(k,l-1)}} g_i(x)$. In particular, we have $\epsilon_i^{(k,l)} = 1 (i = 0, 1)$ in homogeneous networks where all groups are statistically identical; in such circumstances, the derivations of (17) and (18) are simplified to (13) and (14) respectively.

Till now, we have obtained the mean and peakedness of the overflow traffic from the equivalent system for tier $l-1$ network. These results contain the effect of the multi-rate Hayward’s approximation, because the equivalent system is offered with the approximated Poisson traffic $(\Lambda_i^{(k,l-1)}/Z_i^{(k,l-1)}, 1, d_k Z_i^{(k,l-1)})$ obtained by the multi-rate Hayward’s approximation. This effect should be eliminated when considering the mean and peakedness of the overflow traffic from the original tier $l-1$ network. This problem can be solved by the aforementioned semi-equivalent loss model in Section III. Based on (9) and (11), the mean and peakedness of class $k$ overflow traffic from the original tier $l-1$ network to tier $l$ network, denoted as $m_i^{(k,l)}$ and $\tilde{z}_i^{(k,l)} (i = 0, 1)$, are obtained as

$$m_i^{(k,l)} = \tilde{m}_i^{(k,l)} Z_i^{(k,l-1)} = \frac{B_i^{(k,l-1)} \Lambda_i^{(k,l-1)}}{\epsilon_i^{(k,l)}},$$  

(20)

$$\tilde{z}_i^{(k,l)} = \tilde{z}_i^{(k,l)} Z_i^{(k,l-1)}.$$  

(21)

The obtained $m_i^{(k,l)} (i = 0, 1)$ also represents the offered load intensity of the overflow traffic from tier $l-1$ network to tier $l$ network, that is $m_i^{(k,l)} = m_i^{(k,l)}$ at tier $l$ network. Remind that we assume each tier $l$ network group overlaying $J$ tier $l-1$ network groups which are independent and identical. Thus, the upward overflow traffic of a specific class to tier $l$ network group can be obtained as the additive superposition of the individual upward overflow traffic from the overlaid $J$ groups at tier $l-1$. Now including the downward overflow traffic from tier $l+1$ network groups, we have the mean and variance of class $k$ traffic input to tier $l$ network group, denoted as $\Lambda_i^{(k,l)}$ and $\tilde{v}_i^{(k,l)}$, obtained by

$$\Lambda_i^{(k,l)} = a_i^{(k,l)} + J \rho_i^{(k,l)} + \omega_i^{(k,l+1)},$$  

(22)

$$\tilde{v}_i^{(k,l)} = a_i^{(k,l)} + J \rho_i^{(k,l)} \tilde{z}_i^{(k,l)} + \omega_i^{(k,l+1)} \tilde{z}_i^{(k,l+1)}.$$  

(23)

Then the peakedness of class $k$ traffic input to tier $l$ network group is obtained as

$$\tilde{Z}_i^{(k,l)} = \frac{\tilde{V}_i^{(k,l)}}{\Lambda_i^{(k,l)}} = a_i^{(k,l)} + J \rho_i^{(k,l)} \tilde{z}_i^{(k,l)} + \omega_i^{(k,l+1)} \tilde{z}_i^{(k,l+1)}.$$  

(24)

V. PERFORMANCE EVALUATION

Without intensive computation, there are no exact solutions for loss analysis in either homogeneous or heterogeneous hierarchical multiservice systems. We verify our proposed approximation loss model by comparing the analytical results with the results obtained by simulations using OPNET packages [15]. The performance of a three-tier heterogeneous overlay cellular system is evaluated based on the following parameter settings. The link capacity of the network groups from the bottom tier (tier 1) to the top tier (tier 3) is 10 channels, 20 channels and 40 channels, respectively. Each channel is assumed to transmit at a basic rate 16 Kbps. The network groups are assumed to be circular cells. Each tier $l$ network group overlays seven tier $l-1$ network groups. We use the commonly adopted fluid-flow mobility model [16] to determine the mean residence time of a mobile user in a cell. The mean residence time for a call in a tier $l$ group is defined as $(\pi r_l^2)/(2V)$, where $V$ is the mean velocity of the mobile user; $r_l$ is the radius of tier $l$ cell. Two types of mobile users are assumed, slow speed users with mean velocity 1.0 km/hour and fast speed users with mean velocity 50.0 km/hour. The radius of the cells from the bottom to the top tier is 100 m, 10 km and 50 km, respectively. Two classes of source call traffic are assumed to each network group. The traffic of class 1 is transmitted at the basic rate of 16 Kbps and the traffic of class 2 is transmitted at a rate of 32 Kbps. The blocked calls in the bottom tier is overflowed to the upper tiers. The mean call inter-arrival time of the source traffic is from 10 s to 100 s for class 1 and 2 respectively. The mean call holding time is 60 s for both classes. We evaluate the call-level loss performance, as well as the impact of mobility on traffic overflow by assuming each class of source traffic with different hybrid patterns of slow-speed and fast-speed calls. The hybrid pattern for a specific class of source traffic is defined as the proportions of different speed users in the arriving new calls within a unit time, e.g., 20% of new calls
within a unit time is from slow-speed users and 80% is from fast-speed users. We compare the system performance under the following source traffic hybrid patterns:

- Pattern 1 – 20% slow calls and 80% fast calls in class 1, 80% slow calls and 20% fast calls in class 2.
- Pattern 2 – 50% slow calls and 50% fast calls in class 1, 50% slow calls and 50% fast calls in class 2.
- Pattern 3 – 100% slow calls and 0 fast call in class 1, 0 slow call and 100% fast calls in class 2.

In addition, capacity reservation is used to protect handoff calls. To achieve balanced call blocking probability for class 1 and 2 services, we assume that in each tier group two channels are reserved for class 2 handoff calls and one channel reserved for class 1 handoff calls. In simulation, we assign adequate number of cells in each tier to eliminate the effect of mobile users in edge cells.

Fig. 2 and Fig. 3 present the final new call blocking probabilities for class 1 and 2 source traffic of the bottom tier observed in the bottom as well as the higher tiers. The final new call blocking probability for class \( k \) source traffic observed in tier \( L \geq 1 \) is calculated by \( \prod_{l=1}^{L} B_{k,l} \). Here \( B_{k,l} \) is the local new call blocking probability for class \( k \) calls in tier \( l \) network. Although the plotted analytical and simulation results in the figures are very close to each other, we have the following observations. As the proportion of slow-speed new calls in class 1 source traffic increases from 20% to 100%, the call blocking probability in the one-tier system decreases, but increases in the two-tier and three-tier systems. Similar phenomenon is observed in the performance of call dropping probabilities. The reason is clarified in the following.

Fig. 4 and Fig. 5 show the call dropping probability. In the bottom tier, slow-speed calls of class 2 source traffic require more capacity than class 1 calls, as well as longer service time than class 2 fast-speed calls since slow-speed users have longer residence time, hence we observe reduced handoff dropping probability as the proportion of slow-speed calls in class 2 source traffic decreases, i.e. the corresponding proportion of slow-speed calls in class 1 source traffic increases from 20% to 100% as shown in the figures.

However, an opposite effect is observed at the higher tiers. Fast-speed calls at the lower tier with smaller coverage experience more frequent handoffs between the neighboring cells than slow-speed calls. It incurs higher possibility of handoff failure for fast-speed calls, as shown in Fig. 4 and Fig. 5. In the two-tier and three-tier scenarios, the traffic blocked at the lower tiers (tier 1 or tier 2) is overflowed upward to the higher tiers and share the bandwidth with the local calls of the higher tier networks. In our example, the increasing proportion of slow-speed calls in class 1 source traffic corresponds to the increasing proportion of fast-speed calls in class 2 source traffic. This results increased overflows of fast-speed calls from the lower tier to the higher tier, and correspondingly increased new call blocking probability for both classes and increased call dropping probability for class 2 fast-speed calls at higher tiers, as shown in the figures by both analysis and simulation. We also observe that a large proportion of overflow traffic from the lower tier to the higher tier is due to the overflows of fast-speed calls.

VI. CONCLUSIONS

We have proposed an approximate performance modeling for traffic loss analysis in hierarchical heterogeneous wireless networks by taking into account of the statistical characteristics of input traffic, mobility models and service time distribution at different networks. The approximate loss performance obtained by the proposed model are verified by simulation in a hierarchical multi-tier structure. The proposed loss analysis with reduced computation complexity allows efficient network dimensioning for multiservice hierarchical heterogeneous wireless networks.

2There are no data presented in the figures for class 1 fast-speed calls and class 2 slow-speed calls for hybrid Pattern 3.
The application of the approximate performance model proposed in this paper is limited by the assumption that the networks in the same tier are homogeneous with identical statistical characteristics. Future work will focus on the case of heterogeneous networks coexisted in the same tier.

REFERENCES


