# Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights

Sarabjot Singh, *Student Member, IEEE,* Harpreet S. Dhillon, *Student Member, IEEE,* and Jeffrey G. Andrews, *Fellow, IEEE*

***Abstract*—Pushing data traffic from cellular to WiFi is an example of inter radio access technology (RAT) offloading. While this clearly alleviates congestion on the over-loaded cellular network, the ultimate potential of such offloading and its effect on overall system performance is not well understood. To address this, we develop a general and tractable model that consists of $M$ different RATs, each deploying up to $K$ different tiers of access points (APs), where each tier differs in transmit power, path loss exponent, deployment density and bandwidth. Each class of APs is modeled as an independent Poisson point process (PPP), with mobile user locations modeled as another independent PPP, all channels further consisting of i.i.d. Rayleigh fading. The distribution of rate over the entire network is then derived for a weighted association strategy, where such weights can be tuned to optimize a particular objective. We show that the optimum fraction of traffic offloaded to maximize SINR coverage is not in general the same as the one that maximizes rate coverage, defined as the fraction of users achieving a given rate.**

***Index Terms*—WiFi offloading, heterogeneous cellular networks, rate coverage, weighted Poisson Voronoi, stochastic geometry.**

## I. Introduction

WIRELESS networks are facing explosive data demands driven largely by video. While operators continue to rely on their (macro) cellular networks to provide wide-area coverage, they are eager to find complementary alternatives to ease the pressure, especially in areas where subscriber density is high. Complementing the fast evolving heterogeneous cellular networks (HCNs) [2] with the already widely deployed WiFi APs is very attractive to operators and a key aspect of their strategy [3]. In fact, WiFi access points (APs) along with femtocells are projected to carry over 60% of all the global data traffic by 2015 [4]. A future wireless heterogeneous network (HetNet) can be envisioned to have operator-deployed macro base stations (BSs) providing a coverage blanket, along with pico BSs, low powered user-deployed femtocells and user/operator-deployed low powered WiFi APs.

Manuscript received August 13, 2012; revised November 20, 2012 and January 14, 2013; accepted January 19, 2013. The associate editor coordinating the review of this paper and approving it for publication was S. Liew.

This work has been supported by the Intel-Cisco Video Aware Wireless Networks (VAWN) program and NSF grant CIF-1016649. A part of this paper is accepted for presentation at IEEE ICC 2013 in Budapest, Hungary [1].

The authors are with the Wireless Networking and Communications Group (WNCG), The University of Texas at Austin (e-mail: {sarabjot, dhillon}@utexas.edu, jandrews@ece.utexas.edu).

Digital Object Identifier 10.1109/TWC.2013.040413.121174

### A. Motivation and Related Work

Aggressively offloading mobile users from macro BSs to smaller BSs like WiFi hotspots, however, can lead to degradation of user-specific as well as network wide performance. For example, a WiFi AP with excellent signal strength may suffer from heavy load or have less effective bandwidth (channels), thus reducing the effective rate it can serve at [5]. On the other hand, a conservative approach may result in load disparity, which not only leads to underutilization of resources but also degrades the performance of multimedia applications due to bursty interference caused by the lightly loaded APs [6]. Clearly, in such cases any offloading strategy agnostic to these conditions is undesirable, which emphasizes the importance of more adaptive offloading strategies.

RAT selection has been studied extensively in earlier works both from centralized as well as decentralized aspects (see [7] for a survey). Fully centralized schemes, such as in [8]–[10], try to maximize a network wide utility as a solution to the association optimization problem. Decentralized schemes have been studied from game theoretic approaches in [11]–[13] and as heuristic randomized algorithms in [14]. However most of these works focused on flow level assignment and lacked explicit spatial location modeling of the APs and users and the corresponding impact on association. The presented work is more similar to "cell breathing" [15]–[17], wherein the BS association regions are expanded or shrunk depending on the load. Contemporary cellular standards like LTE use a cell breathing approach to address the problem of load balancing in HCNs through cell range expansion (CRE) [2], [18] where users are offloaded to smaller cells using an association bias. A positive association bias implies that a user would be offloaded to a smaller BS as soon as the received power difference from the macro and small BS drops below the bias value. The presented work employs CRE to tune the aggressiveness of offloading from one RAT to another in HetNets. Tools from Poisson point process (PPP) theory and stochastic geometry [19] allow us to quantify the optimal association bias of each constituent tier of each RAT, which maximizes the fraction of time a typical user in the network is served with a rate greater than its minimum rate requirement.

The metric of rate coverage used in this paper, which signifies the fraction of user population able to meet their rate thresholds, captures the inelasticity of traffic such as video services [20], whereas traditional utility based metrics are more suitable for elastic traffic with no hard rate thresholds. There has been considerable advancement in the theory of

HCNs [21]–[23] whereby the locations of APs of each tier are assumed to form a homogeneous PPP. The case of modeling macro cellular networks using a PPP has been strengthened through empirical validation in [24] and theoretical validation in [25]. Load distribution was derived for macro cellular networks in [26] and an empirical fitting based approach was proposed in [27] for association area distribution in a two-tier cellular network. See [6], [28], [29] for a spectral efficiency analysis, where load is modeled through activity of AP queues. While the PPP assumption offers attractive tractability in modeling interference and hence the signal-to-interference-and-noise ratio (SINR) in HetNets, the distribution of *rate* has been elusive. Superposition of point processes, each denoting a class[1] of APs, leads to the formation of disparate association regions (and hence load distribution) due to the unequal transmit powers, path loss exponents and association weights among different classes of APs. Thus, resolving to complicated system level simulations for investigating impact of various wireless algorithms on rate, even for preliminary insights, is not uncommon. One of the goals of this paper is to bridge this gap and provide a tractable framework for deriving the rate distribution in HetNets.

### B. Contributions

The contributions of this paper can be categorized under two main headings.

1) **Modeling and Analysis.** A general $M$-RAT $K$-tier HetNet model is proposed with each class of APs drawn from a homogeneous PPP. This is similar to [21]–[23] with the key difference being the APs of a RAT act as interferers to only the user associated with that RAT. For example, cellular BSs do not interfere with the users associated with a WiFi AP and vice versa. The proposed model is validated by comparing the analytical results with those of a realistic multi-RAT deployment in Section III-E.
   ***Association Regions in HetNet:*** Based on the weighted path loss based association used in this work, the tessellation formed by association regions of APs (region served by the AP) is characterized as a general form of the multiplicatively weighted Poisson Voronoi (PV). Much progress has been made in modeling the area of Poisson Voronoi, see [30]–[32] and references therein, however that of a general multiplicatively weighted PV is an open problem. We propose an analytic approximation for characterizing the association areas (and hence the load) of an AP, which is shown to be quite accurate in the context of rate coverage.
   ***Rate Distribution in HetNet:*** We derive the rate complementary cumulative distribution function (CCDF) of a typical user in the presented HetNet setting in Section III. Rate distribution incorporates congestion in addition to the proximity effects that may not be accurately captured by the SINR distribution alone. Under certain plausible scenarios the derived expression is in closed form and provides insight into system design.

[1]A class refers to a distinct RAT-tier pair.

2) **System Design Insights.** This work allows the inter-RAT offloading to be seen through the prism of association bias wherein the bias can be tuned to suit a network wide objective. We present the following insights in Section IV and V.
   SINR ***Coverage:*** The probability that a randomly located user has SINR greater than an arbitrary threshold is called SINR coverage; equivalently this is the CCDF of SINR. In a simplified two-RAT scenario, e.g. cellular and WiFi, it is shown that the optimal amount of traffic to be offloaded, from one to another, depends solely on their respective SINR thresholds. The optimal association bias, however, is shown to be inversely proportional to the density and transmit power of the corresponding RAT. The maximum SINR coverage under the optimal association bias is then shown to be independent of the density of APs in the network.
   ***Rate Coverage:*** The probability that a randomly located user has rate greater than an arbitrary threshold is called rate coverage; equivalently this is the CCDF of rate. We show that the amount of traffic to be routed through a RAT for maximizing rate coverage can be found analytically and depends on the ratio of the respective resources/bandwidth at each RAT and the user's respective rate (QoS) requirements. Specifically, higher the corresponding ratio, the more traffic should be routed through the corresponding RAT. Also, unlike SINR coverage, the optimal traffic offload fraction increases with the density of the corresponding RAT. Further, the rate coverage always increases with the density of the infrastructure.

## II. System Model

The system model in this paper considers up to a $K$-tier deployment of the APs for each of the $M$-RATs. The set of APs belonging to the same RAT operate in the same spectrum and hence do not interfere with the APs of other RATs. The locations of the APs of the $k^{\text{th}}$ tier of the $m^{\text{th}}$ RAT are modeled as a 2-D homogeneous PPP, $\Phi_{mk}$, of density (intensity) $\lambda_{mk}$. Also, for every class $(m,k)$ there might be BSs allowing no access (closed access) and thus acting only as interferers. For example, subscribers of a particular operator are not able to connect to another operator's WiFi APs but receive interference from them. Such closed access APs are modeled as an independent tier ($k'$) with PPP $\Phi_{mk'}$ of density $\lambda_{mk'}$. The set of all such pairs with non-zero densities in the network is denoted by $\mathcal{V} \triangleq \bigcup_{m=1}^{M} \bigcup_{k\in\mathcal{V}_m} (m,k)$ with $\mathcal{V}_m$ denoting the set of all the tiers of RAT-$m$, i.e., $\mathcal{V}_m = \{k : \lambda_{mk} + \lambda_{mk'} \neq 0\}$. Similarly, $\mathcal{V}_m^o$ and $\mathcal{V}_m^c$ is used to denote the set of open and closed access tiers of RAT-$m$, respectively. Further, the set of open access classes of APs is $\mathcal{V}^o \triangleq \bigcup_{m=1}^{M} \bigcup_{k\in\mathcal{V}_m^o} (m,k)$. The users in the network are assumed to be distributed according to an independent homogeneous PPP $\Phi_u$ with density $\lambda_u$.

Every AP of $(m,k)$ transmits with the same transmit power $\mathrm{P}_{mk}$ over bandwidth $\mathrm{W}_{mk}$. The downlink desired and interference signals are assumed to experience path loss with a path loss exponent $\alpha_k$ for the corresponding tier $k$. The power received at a user from an AP of $(m,k)$ at a distance $x$

TABLE I: Notation Summary

| Notation | Description |
|---|---|
| $M$ | Maximum number of RATs in the network |
| $K$ | Maximum number of tiers of a RAT |
| $(m,k)$ | Pair denoting the $k^{\text{th}}$ tier of the $m^{\text{th}}$ RAT |
| $\mathcal{V};\mathcal{V}^o$ | The set of classes of APs $\bigcup_{m=1}^{M}\bigcup_{k\in\mathcal{V}_m}(m,k)$, where $\mathcal{V}_m=\{k:\lambda_{mk}+\lambda_{mk'}\neq 0\}$; the set of open access classes of APs $\bigcup_{m=1}^{M}\bigcup_{k\in\mathcal{V}_m^o}(m,k)$, where $\mathcal{V}_m^o=\{k:\lambda_{mk}\neq 0\}$ |
| $\Phi_{mk};\Phi_{mk'};\Phi_u$ | PPP of the open access APs of $(m,k)$; PPP of the closed access APs of $(m,k)$; PPP of the mobile users |
| $\lambda_{mk};\lambda_{mk'};\lambda_u$ | Density of open access APs of $(m,k)$; density of closed access APs of $(m,k)$; density of mobile users |
| $\mathrm{T}_{mk};\hat{\mathrm{T}}_{mk}$ | Association weight for $(m,k)$; normalized (divided by that of the serving AP) association weight for $(m,k)$ |
| $\mathrm{P}_{mk};\hat{\mathrm{P}}_{mk}$ | Transmit power of APs of $(m,k)$, specifically $\mathrm{P}_{m1}=53$ dBm, $\mathrm{P}_{m2}=33$ dBm, $\mathrm{P}_{m3}=23$ dBm; normalized transmit power of APs of $(m,k)$ |
| $\mathrm{B}_{mk};\hat{\mathrm{B}}_{mk}$ | Association bias for $(m,k)$; normalized association bias for $(m,k)$. |
| $\alpha_k;\hat{\alpha}_k$ | Path loss exponent of $k^{\text{th}}$ tier; normalized path loss exponent of $k^{\text{th}}$ tier |
| $\sigma_m^2$ | Thermal noise power corresponding to $m^{\text{th}}$ RAT |
| $\mathrm{W}_{mk}$ | Effective bandwidth at an AP of $(m,k)$ |
| $\tau_{mk}$ | SINR threshold of user when associated with $(m,k)$ |
| $\rho_{mk}$ | Rate threshold of user when associated with $(m,k)$ |
| $N_{mk}$ | Load (number of users) associated with an AP of $(m,k)$ |
| $C_{mk}$ | Association area of a typical AP of $(m,k)$ |
| $\mathcal{S}_{mk};\mathcal{S}$ | SINR coverage of user when associated with $(m,k)$; overall SINR coverage of user |
| $\mathcal{R}_{mk};\mathcal{R}$ | Rate coverage of user when associated with $(m,k)$; overall rate coverage of user |

is $\mathrm{P}_{mk}hx^{-\alpha_k}$ where $h$ is the channel power gain. The random channel gains are Rayleigh distributed with average power of unity, i.e., $h\sim\exp(1)$. The general fading distributions can be considered at some loss of tractability [33]. The noise is assumed additive with power $\sigma_m^2$ corresponding to the $m^{\text{th}}$ RAT. Readers can refer to Table I for quick access to the notation used in this paper. In the table and the rest of the paper, the normalized value of a parameter of a class is its value divided by the value it takes for the class of the serving AP.

### A. User Association

For the analysis that follows, let $Z_{mk}$ denote the distance of a typical user from the nearest AP of $(m,k)$. In this paper, a general association metric is used in which a mobile user is connected to a particular RAT-tier pair $(i,j)$ if

$$(i,j)=\arg\max_{(m,k)\in\mathcal{V}^o}\mathrm{T}_{mk}Z_{mk}^{-\alpha_k}, \tag{1}$$

where $\mathrm{T}_{mk}$ is the association weight for $(m,k)$ and ties are broken arbitrarily. These association weights can be tuned to suit a certain network-wide objective. As an example, if $\mathrm{T}_{1k}\gg\mathrm{T}_{2k}$, then more traffic is routed through RAT-1 as compared to RAT-2. Special cases for the choice of association weights, $\mathrm{T}_{mk}$, include:

- $\mathrm{T}_{mk}=1$: the association is to the nearest base station.
- $\mathrm{T}_{mk}=\mathrm{P}_{mk}\mathrm{B}_{mk}$: is the cell range expansion (CRE) technique [2] wherein the association is based on the maximum biased received power, with $\mathrm{B}_{mk}$ denoting the association bias corresponding to $(m,k)$.
- Further, if $\mathrm{B}_{mk}\equiv 1$, then the association is based on maximum received power.

Note that "$\equiv$" is henceforth used to assign the same value to a parameter for all classes of APs, i.e., $x_{mk}\equiv c$ is equivalent to $x_{mk}=c\ \forall\ (m,k)\in\mathcal{V}$. The optimal association weights maximizing rate coverage would depend on load, SINR, transmit powers, densities, respective bandwidths, and path loss exponents of AP classes in the network. Further discussion on the design of optimal association weights is deferred to Section IV. For notational brevity the normalized parameters of $(m,k)$, conditioned on $(i,j)$ being the serving class, are

$$\hat{\mathrm{T}}_{mk}\triangleq\frac{\mathrm{T}_{mk}}{\mathrm{T}_{ij}},\ \hat{\mathrm{P}}_{mk}\triangleq\frac{\mathrm{P}_{mk}}{\mathrm{P}_{ij}},\ \hat{\mathrm{B}}_{mk}\triangleq\frac{\mathrm{B}_{mk}}{\mathrm{B}_{ij}}\ ,\ \hat{\alpha}_k\triangleq\frac{\alpha_k}{\alpha_j}.$$

The association model described above leads to the formation of association regions in the Euclidean plane as described below.

**Definition 1. Association region** of an AP is the region of the Euclidean plane in which all users are served by the corresponding AP. Mathematically, the association region of an AP of class $(i,j)$ located at $x$ is

$$\mathcal{C}_{x_{ij}}=\left\{y\in\mathbb{R}^2:\|y-x\|\leq\left(\frac{\mathrm{T}_{ij}}{\mathrm{T}_{mk}}\right)^{1/\alpha_j}\|y-X_{mk}^*(y)\|^{\hat{\alpha}_k}\ \forall\ (m,k)\in\mathcal{V}^o\right\}, \tag{2}$$

where $X_{mk}^*(y)=\arg\min_{x\in\Phi_{mk}}\|y-x\|$.

The readers familiar with the field of spatial tessellations would recognize that the random tessellation formed by the collection $\{\mathcal{C}_{x_{ij}}\}$ of association regions is a general case of the circular Dirichlet tessellation [34]. The circular Dirichlet tessellation (also known as multiplicatively weighted Voronoi) is the special case of the presented model with equal path loss coefficients. Fig. 1 shows the association regions with two classes of APs in the network ($\mathcal{V}=\{(1,1);(2,3)\}$, say) for two ratios of association weights $\frac{\mathrm{T}_{11}}{\mathrm{T}_{23}}=20$ dB and $\frac{\mathrm{T}_{11}}{\mathrm{T}_{23}}=10$ dB. The path loss exponent is $\alpha_k\equiv 3.5$.

### B. Resource Allocation

A *saturated* resource allocation model is assumed in the downlink of all the APs. This assumption implies that each AP always has data to transmit to its associated mobile users and hence users can be allocated more rate than their rate thresholds. Under the assumed resource allocation, each user receives rate proportional to its link's spectral efficiency. Thus, the rate of a user associated with $(i,j)$ is given by

$$R_{ij}=\frac{\mathrm{W}_{ij}}{N_{ij}}\log\left(1+\mathtt{SINR}_{ij}\right), \tag{3}$$

where $N_{ij}$ denotes the total number of users served by the AP, henceforth referred to as the *load*. The presented rate

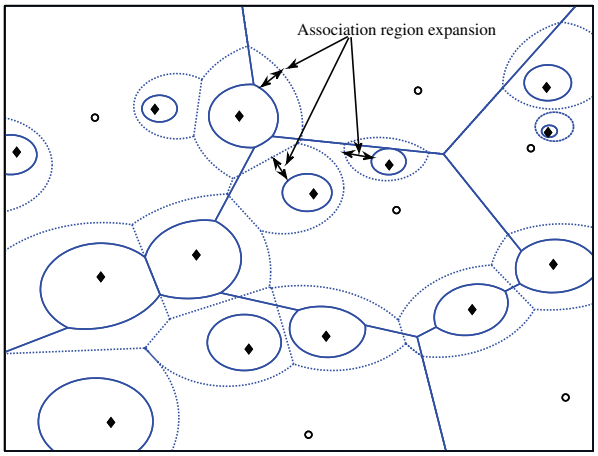



Fig. 1: Association regions of a network with $\mathcal{V} = \{(1,1);(2,3)\}$. The APs of $(1,1)$ are shown as hollow circles and those of $(2,3)$ are shown as solid diamonds. Solid lines show the association regions with $\frac{\mathrm{T}_{11}}{\mathrm{T}_{23}} = 20$ dB and dotted lines show the expanded association regions of $(2,3)$ resulting from the use of $\frac{\mathrm{T}_{11}}{\mathrm{T}_{23}} = 10$ dB.

model captures both the congestion effect (through load) and proximity effect (through SINR). For 4G cellular systems, this rate allocation model has the interpretation of scheduler allocating the OFDMA resources "fairly" among users. For 802.11 CSMA networks, assuming equal channel access probabilities [10], [14] across associated users, leads to the rate model (3). Although the above mentioned resource allocation strategy is assumed in the paper, the ensuing analysis can be extended to a RAT-specific resource allocation methodology as well.

## III. Rate Coverage

This section derives the rate coverage and is the main technical section of the paper. The rate coverage is defined as

$$\mathcal{R} \triangleq \mathbb{P}(R > \rho), \tag{4}$$

and can be thought of equivalently as: (i) the probability that a randomly chosen user can achieve a target rate $\rho$, (ii) the average fraction of users in the network that achieve rate $\rho$, or (iii) the average fraction of the network area that is receiving rate greater than $\rho$.

### A. Load Characterization

This section analyzes the load, which is crucial to get a handle on the rate distribution. The following analysis uses the notion of typicality, which is made rigorous using Palm theory [19, Chapter 4].

**Lemma 1.** *The load at a typical AP of $(i,j)$ has the probability generating function (PGF) given by*

$$\mathsf{G}_{N_{ij}}(z) = \mathbb{E}\left[\exp\left(\lambda_u C_{ij}\left(z-1\right)\right)\right], \tag{5}$$

*where $C_{ij}$ is the association area of a typical AP of $(i,j)$.*

*Proof:* We consider the process $\Phi_{ij} \cup \{0\}$ obtained by adding an AP of $(i,j)$ at the origin of the coordinate system, which is the typical AP under consideration. This is allowed by Slivnyak's theorem [19], which states that the properties observed by a typical[2] point of the PPP, $\Phi_{ij}$, is same as those observed by the point at origin in the process $\Phi_{ij} \cup \{0\}$. The random variable (RV) $N_{ij}$ is the number of users from $\Phi_u$ lying in the association region $\mathcal{C}_{0_{ij}}$ of the typical cell constructed from the process $\Phi_{ij} \cup \{0\}$. Letting $C_{ij}$ denote the random area of this typical association region, the PGF of $N_{ij}$ is given by

$$\mathsf{G}_{N_{ij}}(z) = \mathbb{E}\left[z^{N_{ij}}\right] = \mathbb{E}\left[\exp\left(\lambda_u C_{ij}\left(z-1\right)\right)\right],$$

where the property used is that conditioned on $C_{ij}$, $N_{ij}$ is a Poisson RV with mean $\lambda_u C_{ij}$. ■

As per the association rule (1), the probability that a typical user associates with a particular RAT-tier pair would be directly proportional to the corresponding AP density and association weights. The following lemma identifies the exact relationship.

**Lemma 2.** *The probability that a typical user is associated with $(i,j)$ is given by*

$$\mathcal{A}_{ij} = 2\pi\lambda_{ij}\int_0^\infty z \exp\left(-\pi \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) z^{2/\hat{\alpha}_k}\right)\mathrm{d}z, \tag{6}$$

*where*

$$G_{ij}(m,k) = \lambda_{mk}\hat{\mathrm{T}}_{mk}^{2/\alpha_k}. \tag{7}$$

*If $\alpha_k \equiv \alpha$, then the association probability is simplified to*

$$\mathcal{A}_{ij} = \frac{\lambda_{ij}}{\sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)}. \tag{8}$$

*Proof:* The result can be proved by a minor modification of Lemma 1 of [22]. The proof is presented in Appendix A for completeness. ■

The following two remarks provide alternate interpretations of the association probability.

*Remark* 1. The probability that a typical user is associated with the $i^{\text{th}}$ RAT is given by $\mathcal{A}_i = \sum_{j\in\mathcal{V}_i^o}\mathcal{A}_{ij}$. This probability is also the average fraction of the traffic offloaded, referred henceforth as *traffic offload fraction*, to the $i^{\text{th}}$ RAT.

*Remark* 2. Using the ergodicity of the PPP, $\mathcal{A}_{ij}$ is the average fraction of the total area covered by the association regions of the APs of $(i,j)$.

Based on Remark 2 we note that the mean association area of a typical AP of $(i,j)$ is $\frac{\mathcal{A}_{ij}}{\lambda_{ij}}$. Below we propose a linear scaling based approximation for association areas in HetNets, which matches this first moment. The results based on the area approximation are validated in Section III-E.

**Area Approximation**: The area $C_{ij}$ of a typical AP of the $j^{\text{th}}$ tier of the $i^{\text{th}}$ RAT can be approximated as

$$C_{ij} = C\left(\frac{\lambda_{ij}}{\mathcal{A}_{ij}}\right), \tag{9}$$

[2]The term typical and random are interchangeably used in this paper.

$$\mathcal{S} = \sum_{(i,j)\in\mathcal{V}^o} 2\pi\lambda_{ij} \int_0^\infty y \exp\left( -\frac{\tau_{ij}}{\mathtt{SNR}_{ij}(y)} - \pi \left\{ \sum_{k\in\mathcal{V}_i} D_{ij}(k,\tau_{ij}) y^{2/\hat{\alpha}_k} + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) y^{2/\hat{\alpha}_k} \right\} \right) \mathrm{d}y\,, \tag{19}$$

where $C(y)$ is the area of a typical PV of density $y$ (a scale parameter).

*Remark* 3. The approximation is trivially exact for a single tier, single RAT scenario, i.e., for $\|\mathcal{V}\| = 1$.

*Remark* 4. If $\mathrm{T}_{mk} \equiv \mathrm{T}$ and $\alpha_k \equiv \alpha$, then the approximation is exact. In this case, $\mathcal{A}_{ij} = \frac{\lambda_{ij}}{\sum_{(m,k)\in\mathcal{V}^o}\lambda_{mk}}$ and

$$C\left(\frac{\lambda_{ij}}{\mathcal{A}_{ij}}\right) = C\left(\sum_{(m,k)\in\mathcal{V}^o} \lambda_{mk}\right). \tag{10}$$

With equal association weights and path loss coefficients, the HetNet model becomes the superposition of independent PPPs, which is again a PPP with density equal to the sum of that of the constituents and hence the resulting tessellation is a PV. The right hand side of the above equation is equivalent to a typical association area of a PV with density $\sum_{(m,k)\in\mathcal{V}^o}\lambda_{mk}$.

*Remark* 5. Using the distribution proposed in [32] for $C(y)$, the distribution of $C_{ij}$ is

$$f_{C_{ij}}(c) = \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_{ij}}{\mathcal{A}_{ij}} \left(\frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right)^{2.5} \exp\left(-3.5\frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right), \tag{11}$$

where $\Gamma(x) = \int_0^\infty \exp(-t)t^{x-1}\mathrm{d}t$ is the gamma function.

To characterize the load at the tagged AP (AP serving the typical mobile user) the implicit area biasing needs to be considered and the PGF of the number of *other* – apart from the typical – users ($N_{o,ij}$) associated with the tagged AP needs to be characterized.

**Lemma 3.** *The PGF of the other users associated with the tagged AP of $(i,j)$ is*

$$\mathsf{G}_{N_{o,ij}}(z) = 3.5^{4.5}\left(3.5 + \frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}(1-z)\right)^{-4.5}. \tag{12}$$

*Furthermore, the moments of $N_{o,ij}$ are given by*

$$\mathbb{E}\left[N_{o,ij}^n\right] = \sum_{k=1}^{n} \left(\frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^k \mathsf{S}(n,k)\mathbb{E}\left[C^{k+1}(1)\right], \tag{13}$$

*where $\mathsf{S}(n,k)$ are Stirling numbers of the second kind*[3].

*Proof:* See Appendix B. ∎

The moments of the typical association region of a PV of unit density can be computed numerically and are also available in [30].

### B. SINR Distribution

The SINR of a typical user associated with an AP of $(i,j)$ located at $y$ is

$$\mathtt{SINR}_{ij}(y) = \frac{\mathrm{P}_{ij} h_y y^{-\alpha_j}}{\sum_{k\in\mathcal{V}_i} I_{ik} + \sigma_i^2}, \tag{14}$$

[3]The notation of Stirling numbers given by $\mathsf{S}(n,k)$ should not be confused with that of SINR coverage, $\mathcal{S}$.

where $h_y$ is the channel gain from the tagged AP located at a distance $y$, $I_{ik}$ denotes the interference from the APs of RAT $i$ in the tier $k$. The set of APs contributing to interference are from $\Phi_{ik} \bigcup \Phi_{ik'} \setminus o \,\forall k \in \mathcal{V}_i$, where $o$ denotes the tagged AP from $(i,j)$. Thus

$$I_{ik} = \mathrm{P}_{ik} \sum_{x\in\Phi_{ik}\setminus o} h_x x^{-\alpha_k} + \mathrm{P}_{ik} \sum_{x'\in\Phi_{ik'}} h_{x'} x'^{-\alpha_k}. \tag{15}$$

For a typical user, when associated with $(i,j)$, the probability that the received SINR is greater than a threshold $\tau_{ij}$, or SINR coverage, is

$$\mathcal{S}_{ij}(\tau_{ij}) \triangleq \mathbb{E}_y\left[\mathbb{P}\{\mathtt{SINR}_{ij}(y) > \tau_{ij}\}\right], \tag{16}$$

and the overall SINR coverage is

$$\mathcal{S} = \sum_{(i,j)\in\mathcal{V}^o} \mathcal{S}_{ij}(\tau_{ij})\mathcal{A}_{ij}. \tag{17}$$

Interestingly, the distance of a typical user to the tagged AP in $(i,j)$, $Y_{ij}$, is not only influenced by $\Phi_{ij}$ but also by $\Phi_{mk}\ \forall (m,k) \in \mathcal{V}^o$, as APs of other open access classes also compete to become the serving AP. The distribution of this distance is given by the following lemma.

**Lemma 4.** *The probability distribution function (PDF), $f_{Y_{ij}}(y)$, of the distance $Y_{ij}$ between a typical user and the tagged AP of $(i,j)$ is*

$$f_{Y_{ij}}(y) = \frac{2\pi\lambda_{ij}}{\mathcal{A}_{ij}} y \exp\left\{ -\pi \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) y^{2/\hat{\alpha}_k} \right\}. \tag{18}$$

*Proof:* See Appendix C. ∎

The following lemma gives the SINR CCDF/coverage over the entire network.

**Lemma 5.** *The SINR coverage of a typical user is given by (19) (at the top of page) where*

$$D_{ij}(k,\tau_{ij}) = \hat{\mathrm{P}}_{ik}^{2/\alpha_k}\left\{\lambda_{ik}\mathsf{Z}\left(\tau_{ij},\alpha_k,\hat{\mathrm{T}}_{ik}\hat{\mathrm{P}}_{ik}^{-1}\right) + \lambda_{ik'}\mathsf{Z}(\tau_{ij},\alpha_k,0)\right\},$$

$$G_{ij}(m,k) = \lambda_{mk}\hat{\mathrm{T}}_{mk}^{2/\alpha_k},\ \ \mathsf{Z}(a,b,c) = a^{2/b}\int_{(\frac{c}{a})^{2/b}}^{\infty} \frac{\mathrm{d}u}{1+u^{b/2}},$$

and $\mathtt{SNR}_{ij}(y) = \frac{\mathrm{P}_{ij} y^{-\alpha_j}}{\sigma_i^2}$.

*Proof:* See Appendix D. ∎

The result in Lemma 5 is for the most general case and involves a single numerical integration along with a lookup table for Z. Lemma 5 reduces to the earlier derived SINR coverage expressions in [24] for $M = K = 1$ (single tier, single RAT) and those in [22] for $M = 1$ (single RAT, multiple tiers).

$$\mathcal{R} = \sum_{(i,j)\in\mathcal{V}^o} 2\pi\lambda_{ij} \sum_{n\geq 0} \frac{3.5^{3.5}}{n!} \frac{\Gamma(n+4.5)}{\Gamma(3.5)} \left(\frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^n \left(3.5 + \frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^{-(n+4.5)} \times \int_0^\infty y \exp\left(-\frac{t(\hat{\rho}_{ij}(n+1))}{\mathtt{SNR}_{ij}(y)} - \pi\left\{\sum_{k\in\mathcal{V}_i} D_{ij}(k, t(\hat{\rho}_{ij}(n+1)))y^{2/\hat{\alpha}_k} + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)y^{2/\hat{\alpha}_k}\right\}\right) \mathrm{d}y\,, \quad (20)$$

$$\bar{\mathcal{R}} = \sum_{(i,j)\in\mathcal{V}^o} 2\pi\lambda_{ij} \int_0^\infty y \exp\left(-\frac{t(\hat{\rho}_{ij}\bar{N}_{ij})}{\mathtt{SNR}_{ij}(y)} - \pi\left\{\sum_{k\in\mathcal{V}_i} D_{ij}(k, t(\hat{\rho}_{ij}\bar{N}_{ij}))y^{2/\hat{\alpha}_k} + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)y^{2/\hat{\alpha}_k}\right\}\right) \mathrm{d}y\,, \quad (26)$$

### C. Main Result

Having characterized the distribution of load and SINR, we now derive the rate distribution over the whole network.

**Theorem 1.** *The rate coverage of a randomly located mobile user in the general HetNet setting of Section II is given by (20) (at the top of page) where $\rho_{ij}$ is the rate threshold for $(i,j)$, $\hat{\rho}_{ij} \triangleq \rho_{ij}/\mathrm{W}_{ij}$, and $t(x) \triangleq 2^x - 1$.*

*Proof:* Using (3), the probability that the rate requirement of a user associated with $(i,j)$ is met is

$$\begin{aligned}\mathbb{P}(R_{ij} > \rho_{ij}) &= \mathbb{P}\left(\frac{\mathrm{W}_{ij}}{N_{ij}} \log(1+\mathtt{SINR}_{ij}) > \rho_{ij}\right) \\ &= \mathbb{P}(\mathtt{SINR}_{ij} > 2^{\rho_{ij}N_{ij}/\mathrm{W}_{ij}} - 1) \quad (21)\\ &= \mathbb{E}_{N_{ij}}\left[\mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}N_{ij})\right)\right], \quad (22)\end{aligned}$$

where $t(\hat{\rho}_{ij}N_{ij}) = 2^{\rho_{ij}N_{ij}/\mathrm{W}_{ij}} - 1$ and $N_{ij} = 1 + N_{o,ij}$, i.e., the load at the tagged AP equals the typical user plus the *other* users. Using Lemma 3, (22) is simplified as

$$\begin{aligned}&\mathbb{E}_{N_{ij}}\left[\mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}N_{ij})\right)\right] \\ &= \sum_{n\geq 0} \mathbb{P}(N_{o,ij} = n)\mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}(n+1))\right) \quad (23)\\ &= \sum_{n\geq 0} \frac{3.5^{3.5}}{n!} \frac{\Gamma(n+4.5)}{\Gamma(3.5)} \left(\frac{\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}\right)^n \times \left(3.5 + \frac{\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}\right)^{-(n+4.5)} \mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}(n+1))\right). \quad (24)\end{aligned}$$

Using the law of total probability, the rate coverage is

$$\begin{aligned}\mathcal{R} &= \sum_{(i,j)\in\mathcal{V}^o} \mathcal{A}_{ij}\mathbb{P}(R_{ij} > \rho_{ij}) \\ &= \sum_{(i,j)\in\mathcal{V}^o} \mathcal{A}_{ij} \sum_{n\geq 0} \frac{3.5^{3.5}}{n!} \frac{\Gamma(n+4.5)}{\Gamma(3.5)} \times \left(\frac{\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}\right)^n \left(3.5 + \frac{\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}\right)^{-(n+4.5)} \mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}(n+1))\right). \quad (25)\end{aligned}$$

Using Lemma 5 in the above equation gives the desired result. ∎

The rate distribution expression for the most general setting requires a single numerical integral and use of lookup tables for Z and Γ. Since both the terms $\mathbb{P}(N_{ij} = n)$ and $\mathcal{S}_{ij}\left(t(n)\right)$ decay rapidly for large $n$, the summation over $n$ in Theorem 1 can be accurately approximated as a finite summation to a sufficiently large value, $N_{\max}$. We found $N_{\max} = 4\lambda_u$ to be sufficient for results presented in Section III-E.

### D. Mean Load Approximation

The rate coverage expression can be further simplified (sacrificing accuracy) if the load at each AP of $(i,j)$ is assumed equal to its mean.

**Corollary 1.** *Rate coverage with the mean load approximation is given by (26) (at the top of page), where*

$$\bar{N}_{ij} = \mathbb{E}\left[N_{ij}\right] = 1 + \frac{1.28\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}.$$

*Proof:* Lemma 3 gives the first moment of load as $\mathbb{E}\left[N_{ij}\right] = 1 + \mathbb{E}\left[N_{o,ij}\right] = 1 + \frac{\lambda_u\mathcal{A}_{ij}}{\lambda_{ij}}\mathbb{E}\left[C^2(1)\right]$ where $\mathbb{E}\left[C^2(1)\right] = 1.28$ [30]. Using an approximation for (22) with $\mathbb{E}_{N_{ij}}\left[\mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}N_{ij})\right)\right] \approx \mathcal{S}_{ij}\left(t(\hat{\rho}_{ij}\mathbb{E}\left[N_{ij}\right])\right)$, the simplified rate coverage expression is obtained. ∎

The mean load approximation above simplifies the rate coverage expression by eliminating the summation over $n$. The numerical integral can also be eliminated in certain plausible scenarios given in the following corollary.

**Corollary 2.** *In interference limited scenarios ($\sigma^2 \to 0$) with mean load approximation and with same path loss exponents ($\hat{\alpha}_k \equiv 1$), the rate coverage is*

$$\bar{\mathcal{R}} = \sum_{(i,j)\in\mathcal{V}^o} \frac{\lambda_{ij}}{\sum_{k\in\mathcal{V}_i} D_{ij}(k, t(\hat{\rho}_{ij}\bar{N}_{ij})) + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)}. \quad (27)$$

In the above analysis, rate distribution is presented as a function of association weights. So, in principle, it is possible to find the optimal association weights and hence the optimal fraction of traffic to be offloaded to each RAT so as to maximize the rate coverage. This aspect is studied in a special case of a two-RAT network in Section IV.

### E. Validation

In this section, the emphasis is on validating the area and mean load approximations proposed for rate coverage and on validating the PPP as a suitable AP location model. In all the simulation results, we consider a square window of $20 \times 20$ km$^2$. The AP locations are drawn from a PPP or a real deployment or a square grid depending upon the scenario that is being simulated. The typical user is assumed to be located

at the origin. The serving AP for this user (tagged AP) is determined by (1). The received SINR can now be evaluated as being the ratio of the power received from the serving AP and the sum of the powers received from the rest of the APs as given in (14). The rest of the users are assumed to form a realization of an independent PPP. The serving AP of each user is again determined by (1), which provides the total load on the tagged AP in terms of the number of users it is serving. The rate of the typical user is then computed according to (3). In each Monte-Carlo trial, the user locations, the base station locations, and the channel gains are independently generated. The rate distribution is obtained by simulating $10^5$ Monte-Carlo trials.

In the discussion that follows we use a specific form of the association weight as $\mathrm{T}_{mk} = \mathrm{P}_{mk}\mathrm{B}_{mk}$ corresponding to the biased received power based association [2], where $\mathrm{B}_{mk}$ is the association bias for $(m, k)$. The effective resources at an AP are assumed to be uniformly $\mathrm{W}_{mk} \equiv 10$ MHz and equal rate thresholds are assumed for all classes. Thermal noise is ignored. Also, without any loss of generality the bias of $(1, 1)$ is normalized to 1, or $\mathrm{B}_{11} = 0$ dB.

*1) Analysis:* Our goal here is to validate the area approximation and the mean load approximation (Theorem 1 and Corollary 1, respectively) in the context of rate coverage. A scenario with two-RATs, one with a single open access tier and the other with two tiers – one open and one closed access – is considered first. In this case, $\mathcal{V} = \{(1, 1); (2, 3); (2, 3^{'})\}$, $\lambda_{11} = 1$ BS/km$^2$, $\lambda_{23} = \lambda_{23'} = 10$ BS/km$^2$, $\lambda_u = 50$ users/km$^2$, $\alpha_1 = 3.5$, and $\alpha_3 = 4$. Fig. 2 shows the rate distribution obtained through simulation and that from Theorem 1 and Corollary 1 for two values of association biases. Fig. 3 shows the the rate distribution in a two-RAT three-tier setting with $\mathcal{V} = \{(1, 1); (1, 2); (2, 2); (2, 3)\}$, $\lambda_{11} = 1$ BS/km$^2$, $\lambda_{12} = \lambda_{22} = 5$ BS/km$^2$, $\lambda_{23} = 10$ BS/km$^2$, $\lambda_u = 50$ users/km$^2$, $\alpha_1 = 3.5$, $\alpha_2 = 3.8$, and $\alpha_3 = 4$ for two values of association bias of $(2, 3)$. In both cases, $\mathrm{B}_{12} = \mathrm{B}_{22} = 5$ dB.

As it can be observed from both the plots, the analytic distributions obtained from Theorem 1 and Corollary 1 are in quite good agreement with the simulated one and thus validate the analysis. See [1] for validation of a three-RAT scenario.

*2) Spatial Location Model:* To simulate a realistic spatial location model for a two-RAT scenario, the cellular BS location data of a major metropolitan city used in [24] is overlaid with that of an actual WiFi deployment [35]. Along with the PPP, a square grid based location model in which the APs for both the RATs are located in a square lattice (with different densities) is also used in the following comparison. Denoting the macro tier as $(1, 1)$ and WiFi APs as $(2, 3)$, $\mathcal{V} = \{(1, 1); (2, 3)\}$ in this setup. The superposition is done such that $\lambda_{23} = 10\lambda_{11}$. Fig. 4 shows the rate distribution of a typical user obtained from the real data along with that of a square grid based model and that from a PPP, Theorem 1, for three cases. As evident from the plot, Theorem 1 is quite accurate in the context of rate distribution with regards to the actual location data.

## IV. Design of Optimal Offload

In this section, we consider the design of optimal offloading under a specific form of the association weight as $\mathrm{T}_{mk} = \mathrm{P}_{mk}\mathrm{B}_{mk}$. For general settings, the optimum association biases $\{\mathrm{B}_{mk}\}$ for SINR and rate coverage can be found using the derived expressions of Lemma 5 and Theorem 1 respectively. As discussed in Section III-E, simplified expression of Corollary 1 can also be used for rate coverage. We consider below a two-RAT single tier scenario with $q^{\text{th}}$ tier of RAT-1 overlaid with $r^{\text{th}}$ tier of RAT-2, i.e., $\mathcal{V} = \{(1, q); (2, r)\}$. Optimal association bias and optimal traffic offload fraction is investigated here in the context of both the SIR coverage (i.e., neglecting noise) and rate coverage.

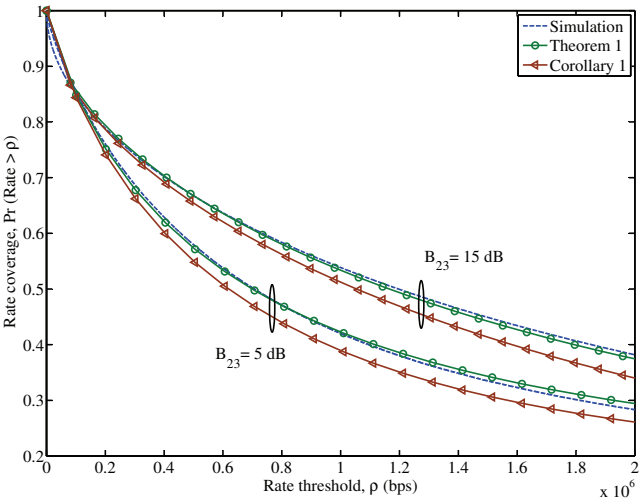



Fig. 2: Comparison of rate distribution obtained from simulation, Theorem 1, and Corollary 1 for $\lambda_{23} = \lambda_{23'} = 10\lambda_{11}$, $\alpha_1 = 3.5$, and $\alpha_3 = 4$.

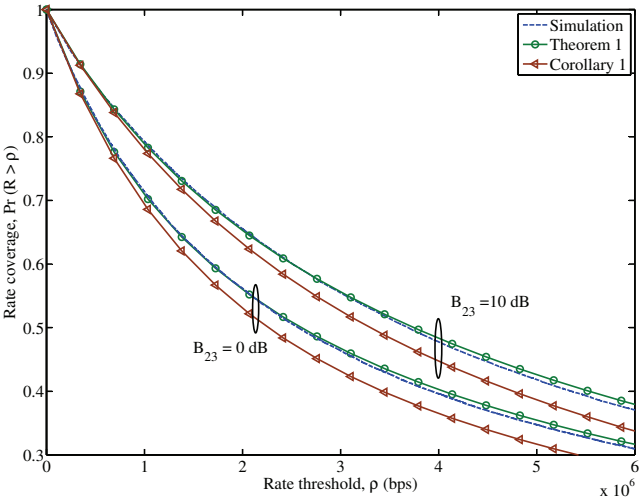



Fig. 3: Comparison of rate distribution obtained from simulation, Theorem 1, and Corollary 1 for $\lambda_{12} = \lambda_{22} = 5\lambda_{11}$, $\lambda_{23} = 10\lambda_{11}$, $\alpha_1 = 3.5$, $\alpha_2 = 3.8$, and $\alpha_3 = 4$.

### A. Offloading for Optimal SIR Coverage

**Proposition 1.** *Ignoring thermal noise (interference limited scenario, $\sigma^2 \to 0$), assuming equal path loss coefficients ($\hat{\alpha}_k \equiv 1$), the value of association bias $\frac{\mathrm{B}_{2r}}{\mathrm{B}_{1q}}$ maximizing* SIR

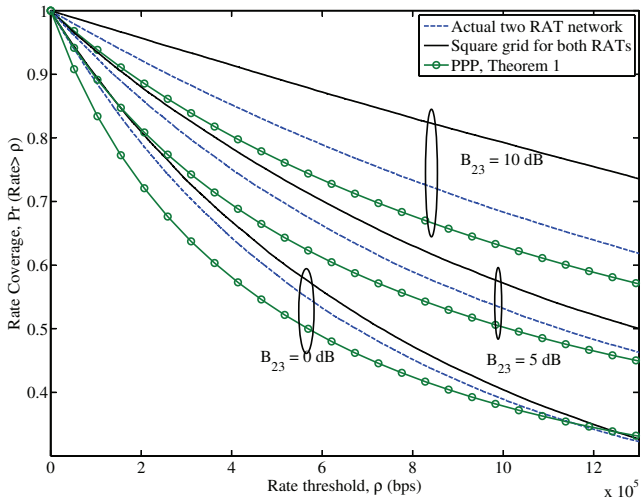



Fig. 4: Rate distribution comparison for the three spatial location models: real, grid, and PPP for a two-RAT setting with $\lambda_{23} = 10\lambda_{11}$ and $\alpha_1 = \alpha_3 = 4$

*coverage is*

$$b_{\text{opt}} = \frac{\mathrm{P}_{1q}}{\mathrm{P}_{2r}} \left( \frac{\mathsf{Z}(\tau_{1q}, \alpha, 1)}{a\mathsf{Z}(\tau_{2r}, \alpha, 1)} \right)^{\alpha/2}, \tag{28}$$

*where* $\lambda_{2r} = a\lambda_{1q}$ *and the corresponding optimum traffic offload fraction to RAT-2 is*

$$\mathcal{A}_2 = \frac{\mathsf{Z}(\tau_{1q}, \alpha, 1)}{\mathsf{Z}(\tau_{2r}, \alpha, 1) + \mathsf{Z}(\tau_{1q}, \alpha, 1)}. \tag{29}$$

*The corresponding* SIR *coverage is*

$$\frac{\mathsf{Z}(\tau_{2r}, \alpha, 1) + \mathsf{Z}(\tau_{1q}, \alpha, 1)}{\mathsf{Z}(\tau_{2r}, \alpha, 1) + \mathsf{Z}(\tau_{1q}, \alpha, 1) + \mathsf{Z}(\tau_{2r}, \alpha, 1)\mathsf{Z}(\tau_{1q}, \alpha, 1)}. \tag{30}$$

*Proof:* See Appendix E. ∎

The following observations can be made from the above Proposition:

- The optimal bias for SIR coverage is inversely proportional to the density and transmit power of the corresponding RAT. This is because the denser the second RAT and the higher the transmit power of the corresponding APs, the higher the interference experienced by offloaded users leading to a decrease in the optimal bias. Also, with increased density and power, lesser bias is required to offload the same fraction of traffic.
- The optimal fraction of traffic/user population to be offloaded to either RAT for maximizing SIR coverage is *independent* of the density and power and is solely dependent on SIR thresholds. The higher the RAT-1 threshold, $\tau_{1q}$, compared to that of RAT-2 threshold, $\tau_{2r}$, the more percentage of traffic is offloaded to RAT-2 as $\mathsf{Z}$ is a monotonically increasing function of $\tau$. In fact, if $\tau_{1q} = \tau_{2r}$, offloading *half* of the user population maximizes SIR coverage.

### B. Offloading for Optimal Rate Coverage

For the design of optimal offloading for rate coverage, the mean load approximation (Corollary 1) is used.

**Proposition 2.** *Ignoring thermal noise (interference limited scenario,* $\sigma^2 \to 0$*), assuming equal path loss coefficients* ($\hat{\alpha}_k \equiv 1$)*, the value of association bias* $\frac{\mathrm{B}_{2r}}{\mathrm{B}_{1q}}$ *maximizing rate coverage is*

$$b_{\text{opt}} = \arg\max_b \left\{ \left( \mathsf{Z}(t_{1q}(\hat{\rho}_{1q}\bar{N}_{1q}), \alpha, 1) + 1 + a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha} \right)^{-1} + \left( \mathsf{Z}(t_{2r}(\hat{\rho}_{2r}\bar{N}_{2r}), \alpha, 1) + 1 + \frac{1}{a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha}} \right)^{-1} \right\}, \tag{31}$$

*where* $a = \lambda_{2r}/\lambda_{1q}$ *and* $b = \mathrm{B}_{2r}/\mathrm{B}_{1q}$.

*Proof:* The optimum association bias can be found by maximizing the expression obtained from Corollary 2 using $\mathcal{V} = \{(1, q); (2, r)\}$, $\lambda_{2r} = a\lambda_{1q}$, and $\mathrm{B}_{2r} = b\mathrm{B}_{1q}$. ∎

Unfortunately, a closed form expression for the optimal bias is not possible in this case, as the load (and hence the threshold) is dependent on the association bias $b$. However, the optimal association bias, $b_{\text{opt}}$, for the rate coverage can be found out through a linear search using the above Proposition. In a general setting, the computational complexity of finding the optimal biases, however, increases with the number of classes of APs in the network as the dimension of the problem increases. While the exact computational complexity depends upon the choice of optimization algorithm, the proposed analytical approach is clearly less complex than exhaustive simulations by virtue of the easily computable rate coverage expression.

The analysis in this section shows that for a two-RAT scenario, SIR coverage and rate coverage exhibit considerably different behavior. The optimal traffic offload fraction for SIR coverage is independent of the density whereas for rate coverage it is expected to increase because of the decreasing load per AP for the second RAT. For a fixed bias, rate coverage always increases with density, however for a fixed density there is always an optimal traffic offload fraction. These insights might be known to practicing wireless system engineers but here a theoretical analysis makes the observations rigorous.

## V. Results and Discussion

In this section we primarily consider a setting of macro tier of RAT-1 overlaid with a low power tier of RAT-2, i.e., $\mathcal{V} = \{(1, 1); (2, 3)\}$. This setting is similar to the widespread use of WiFi APs to offload the macro cell traffic. In particular, the effect of association bias and traffic offload fraction on SIR and rate coverage is investigated. Thermal noise is ignored in the following results.

### A. SIR coverage

The variation of SIR coverage with the density of RAT-2 APs for different values of association bias is shown in Fig. 5. The path loss exponent used is $\alpha_k \equiv 3.5$ and the respective SIR thresholds are $\tau_{11} = 3$ dB and $\tau_{23} = 6$ dB. It is clear that for any fixed value of association bias, $\mathcal{S}$ is sub-optimal for all values of densities except for the bias value satisfying Proposition 1. Also shown is the optimum SIR coverage (Proposition 1), which is invariant to the density of APs.

Variation of SIR coverage with the association bias is shown in Fig. 6 for different densities of RAT-2 APs. As shown, increasing density of RAT-2 APs decreases the optimal offloading bias. This is due to the corresponding increase in the interference for offloaded users in RAT-2. This insight will also be useful in rate coverage analysis. Again, at all values of association bias, $\mathcal{S}$ is sub-optimal for all density values except for the optimum density, $\lambda_{\text{opt}} = \left(\frac{\mathrm{P}_{1q}}{\mathrm{P}_{2r}\mathrm{B}_{2r}}\right)^{2/\alpha} \frac{\mathrm{Z}(\tau_{1q},\alpha,1)}{\mathrm{Z}(\tau_{2r},\alpha,1)}$.

### B. Rate Coverage

The variation of rate coverage with the density of RAT-2 APs for different values of association bias is shown in Fig. 7 and the variation with the association bias is shown in Fig. 8 for different densities of RAT-2 APs. In these results, the user density $\lambda_u = 200$ users/km$^2$, the rate threshold $\rho_{mk} \equiv 256$ Kbps, the effective bandwidth $\mathrm{W}_{mk} \equiv 10$ MHz, and the path loss exponent is $\alpha_k \equiv 3.5$. As expected, rate coverage increases with increasing AP density because of the decrease in load at each AP. The optimum association bias for rate coverage is obtained by a linear search as in Proposition 2. For all values of association bias, $\mathcal{R}$ is sub-optimal except for the one given in Proposition 2. Fig. 9 shows the effect of association bias on the $5^{\text{th}}$ percentile rate $\rho_{95}$ with $\mathcal{R}|_{\rho_{95}} = 0.95$ (i.e., 95% of the user population receives a rate greater than $\rho_{95}$) for different densities of RAT-2 APs. Comparing Fig. 8 and Fig. 9, it can be seen that the optimal bias is agnostic to rate thresholds. This leads to the design insight that for given network parameters re-optimization is not needed for different rate thresholds. The developed analysis can also be used to find optimal biases for a more general setting. Fig. 10 shows the $5^{\text{th}}$ percentile rate for a setting with $\mathcal{V} = \{(1,1);(1,2);(2,2);(2,3)\}$, $\lambda_{11} = 1$ BS/km$^2$, $\lambda_{12} = \lambda_{12} = 5$ BS/km$^2$, $\mathrm{B}_{12} = \mathrm{B}_{22} = 5$ dB as a function of association bias of $(2,3)$. It can be seen that the choice of association biases can heavily influence rate coverage.

A common observation in Fig. 8-10 is the decrease in the optimal offloading bias with the increase in density of APs of the corresponding RAT. This can be explained by the earlier insight of decreasing optimal bias for SIR coverage with increasing density. However, in contrast to the trend in SIR coverage, the optimum traffic offload fraction increases with increasing density as the corresponding load at each AP of second RAT decreases. These trends are further highlighted in Fig. 11 for the following scenarios:

- Case 1: $\mathrm{W}_{11} = 15$ MHz, $\mathrm{W}_{23} = 5$ MHz, $\rho_{11} = 256$ Kbps, and $\rho_{23} = 512$ Kbps.
- Case 2: $\mathrm{W}_{11} = 5$ MHz, $\mathrm{W}_{23} = 15$ MHz, $\rho_{11} = 512$ Kbps, and $\rho_{23} = 256$ Kbps.

It can be seen that apart from the effect of deployment density, optimum choice of association bias and traffic offload fraction also depends on the ratio of rate threshold ($\rho_{ij}$) to the bandwidth ($\mathrm{W}_{ij}$), or $\hat{\rho}_{ij}$. In particular, larger the ratio of the available resources to the rate threshold more is the tendency to be offloaded to the corresponding RAT.

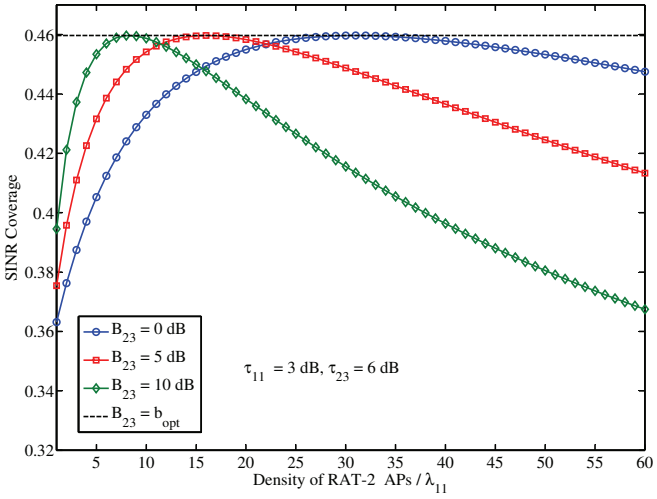



Fig. 5: Effect of density of RAT-2 APs on SINR coverage.

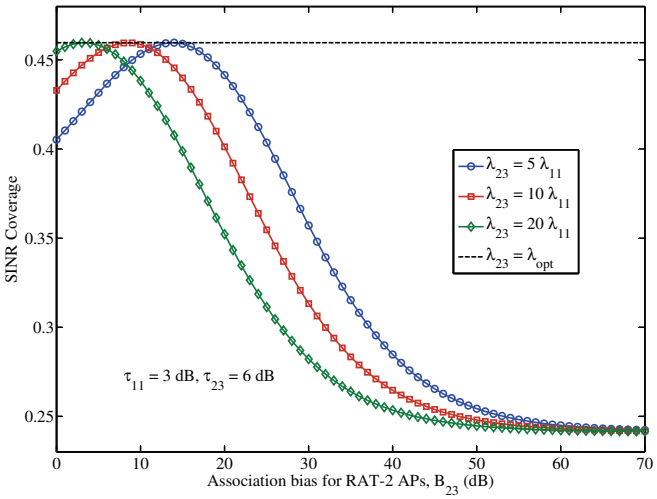



Fig. 6: Effect of association bias for RAT-2 APs on SINR coverage.

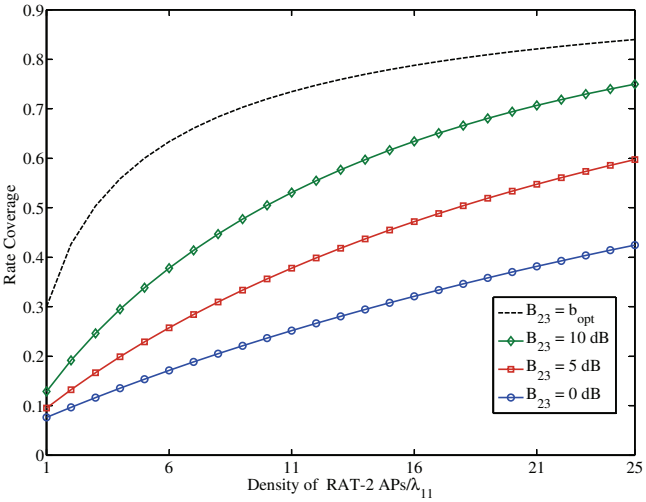



Fig. 7: Effect of density of RAT-2 APs on rate coverage.

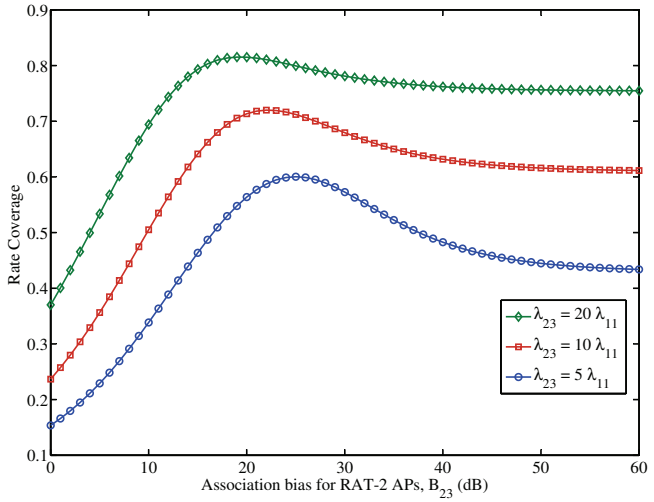



Fig. 8: Effect of association bias for RAT-2 APs on rate coverage.

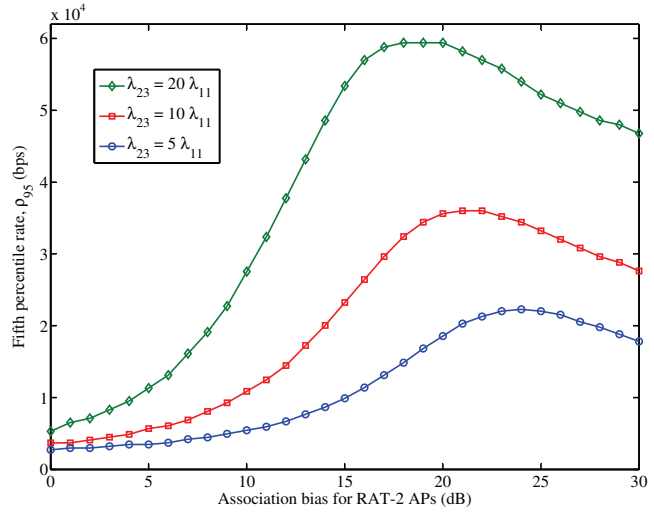



Fig. 9: Effect of association bias for RAT-2 APs on $5^{\text{th}}$ percentile rate with $\mathcal{V} = \{(1,1);(2,3)\}$.

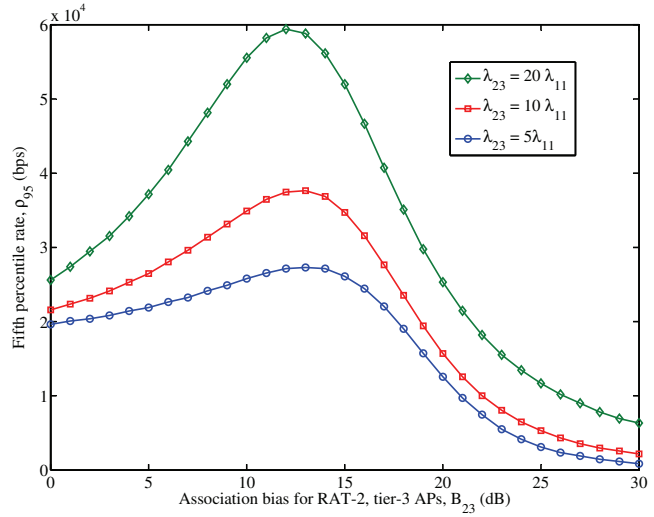



Fig. 10: Effect of association bias for third tier of RAT-2 APs on $5^{\text{th}}$ percentile rate with $\lambda_{12} = \lambda_{22} = 5\lambda_{11}$, $\mathrm{B}_{12} = \mathrm{B}_{22} = 5$ dB.

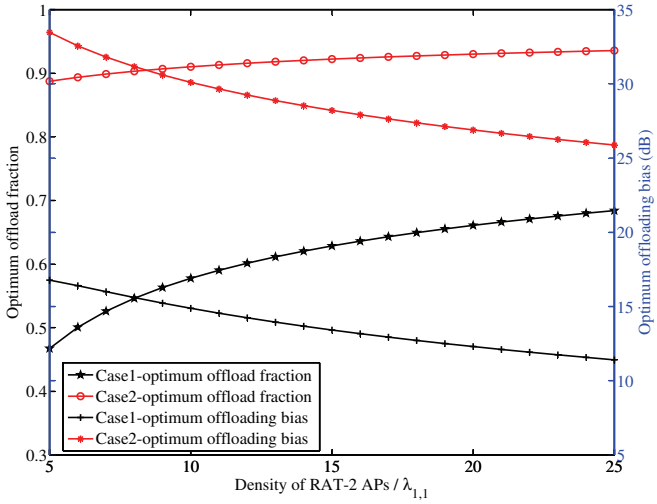



Fig. 11: Effect of user's rate requirements and effective resources on the optimum association bias and optimum traffic offload fraction.

## VI. Conclusion

In this paper, we presented a tractable model to analyze the effects of offloading in a $M$-RAT $K$-tier wireless heterogeneous network setting under a flexible association model. To the best of our knowledge, the presented work is the first to study rate coverage in the context of inter-RAT offload. Using biased received power based association, it is shown that there exists an optimum percentage of the traffic that should be offloaded for maximizing the rate coverage which in turn is dependent on user's QoS requirements and the resource condition at each available RAT besides from the received signal power and load. Investigating the coupling of AP queues induced by offloading, which has been ignored in this work, could be an interesting future extension. Although the emphasis of this work has been on inter-RAT offload, the framework can also be used to provide insights for inter-tier offload within a RAT. Also, the area approximation for the association regions can be improved further by employing a non-linear approximation.

## Appendix A

*Proof of Lemma 2:* If $\mathcal{A}_{ij}$ is the association probability of a typical user with RAT-tier pair $(i,j)$, then

$$\mathcal{A}_{ij} = \mathbb{P}\left( \bigcap_{\substack{(m,k)\in\mathcal{V}^o \\ (m,k)\neq(i,j)}} \left\{ \mathrm{T}_{ij} Z_{ij}^{-\alpha_j} > \mathrm{T}_{mk} Z_{mk}^{-\alpha_k} \right\} \right), \quad (32)$$

since $Z_{mk}$ denotes the distance to nearest AP in $\Phi_{mk}$. Thus

$$\mathcal{A}_{ij} \overset{(a)}{=} \prod_{\substack{(m,k)\in\mathcal{V}^o \\ (m,k)\neq(i,j)}} \mathbb{P}\left( \mathrm{T}_{ij} Z_{ij}^{-\alpha_j} > \mathrm{T}_{mk} Z_{mk}^{-\alpha_k} \right) \quad (33)$$

$$= \int_{z>0} \prod_{\substack{(m,k)\in\mathcal{V}^o \\ (m,k)\neq(i,j)}} \mathbb{P}\left( Z_{mk} > (\hat{\mathrm{T}}_{mk})^{1/\alpha_k} z^{1/\hat{\alpha}_k} \right) f_{Z_{ij}}(z)\mathrm{d}z, \quad (34)$$

where (a) follows from the independence of $\Phi_{mk}$, $\forall (m,k) \in \mathcal{V}$. Now

$$\mathbb{P}\left(Z_{mk} > z\right) = \mathbb{P}\left(\Phi_{mk} \cap b(0,z) = \emptyset\right) = e^{-\pi \lambda_{mk} z^2}, \quad (35)$$

where $b(0,z)$ is the Euclidean ball of radius $z$ centered at origin. The probability distribution function $f_{Z_{mk}}(z)$ can be written as

$$\begin{aligned} f_{Z_{mk}}(z) &= \frac{\mathrm{d}}{\mathrm{d}z}\{1 - \mathbb{P}(Z_{mk} > z)\} \\ &= 2\pi\lambda_{mk} z \exp(-\pi\lambda_{mk} z^2), \ \forall z \geq 0. \end{aligned} \quad (36)$$

Using (34), (35) and (36)

$$\begin{aligned} \mathcal{A}_{ij} &= 2\pi\lambda_{ij} \\ &\times \int_{z>0} z \exp\left( -\pi \sum_{\substack{(m,k)\in\mathcal{V}^o \\ (m,k)\neq(i,j)}} \lambda_{mk} (\hat{\mathrm{T}}_{mk})^{2/\alpha_k} z^{2/\hat{\alpha}_k} \right) \\ &\qquad \times \exp(-\pi\lambda_{ij} z^2)\mathrm{d}z, \end{aligned} \quad (37)$$

which gives (6). ∎

## APPENDIX B

*Proof of Lemma 3:* As a random user is more likely to lie in a larger association region then in a smaller association region, the distribution of the association area of the tagged AP, $C_{ij}^{'}$, is proportional to its area and can be written as

$$f_{C_{ij}^{'}}(c) \propto c f_{C_{ij}}(c). \quad (38)$$

Using the normalization property of the distribution function and (11), the biased area distribution is

$$f_{C_{ij}^{'}}(c) = \frac{c f_{C_{ij}}(c)}{\mathbb{E}\left[C_{ij}\right]} = \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_{ij}}{\mathcal{A}_{ij}} \left(\frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right)^{3.5} \exp\left(-3.5 \frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right). \quad (39)$$

The location of the other users (apart from the typical user) in the association region of the tagged AP follows the reduced Palm distribution of $\Phi_u$ which is the same as the original distribution since $\Phi_u$ is a PPP [19, Sec. 4.4]. Thus, using Lemma 1 and (39), the PGF of the other users in the tagged AP is

$$\begin{aligned} \mathsf{G}_{N_{o,ij}}(z) &= \mathbb{E}\left[\exp\left(\lambda_u C_{ij}^{'}(z-1)\right)\right] \\ &= \int_{c>0} \exp\left(\lambda_u c(z-1)\right) \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_{ij}}{\mathcal{A}_{ij}} \left(\frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right)^{3.5} \\ &\qquad \exp\left(-3.5 \frac{\lambda_{ij}}{\mathcal{A}_{ij}} c\right) \mathrm{d}c \quad (40) \\ &= 3.5^{4.5} \left(3.5 + \frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}(1-z)\right)^{-4.5}. \quad (41) \end{aligned}$$

Using the PGF, the probability mass function can be derived as

$$\begin{aligned} \mathbb{P}\left(N_{ij} = n+1\right) &= \mathbb{P}\left(N_{o,ij} = n\right) = \frac{\mathsf{G}_{N_{o,ij}}^{(n)}(0)}{n!} \\ &= \frac{3.5^{3.5}}{n!} \frac{\Gamma(n+4.5)}{\Gamma(3.5)} \left(\frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^n \times \left(3.5 + \frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^{-(n+4.5)}. \end{aligned} \quad (42)$$

For the second half of the proof, we use the property that the moments of a Poisson RV, $X \sim \mathrm{Pois}(\lambda)$ (say), can be written in terms of Stirling numbers of the second kind, $\mathsf{S}(n,k)$, as $\mathbb{E}\left[X^n\right] = \sum_{k=0}^{n} \lambda^k \mathsf{S}(n,k)$. Now

$$\mathbb{E}\left[N_{o,ij}^n\right] = \mathbb{E}\left[\mathbb{E}\left[N_{o,ij}^n | C_{ij}^{'}\right]\right] \quad (43)$$

$$= \mathbb{E}\left[\sum_{k=0}^{n} (\lambda_u C_{ij}^{'})^k \mathsf{S}(n,k)\right] = \sum_{k=1}^{n} \lambda_u^k \mathsf{S}(n,k) \mathbb{E}\left[C_{ij}^{'k}\right]. \quad (44)$$

Using (39) and the area approximation (9)

$$\mathbb{E}\left[C_{ij}^{'k}\right] = \frac{\mathbb{E}\left[C_{ij}^{k+1}\right]}{\mathbb{E}\left[C_{ij}\right]} = \frac{(\lambda_{ij}/\mathcal{A}_{ij})^{-(k+1)} \mathbb{E}\left[C^{k+1}(1)\right]}{(\lambda_{ij}/\mathcal{A}_{ij})^{-1} \mathbb{E}\left[C(1)\right]}, \quad (45)$$

and thus

$$\mathbb{E}\left[N_{o,ij}^n\right] = \sum_{k=1}^{n} \left(\frac{\lambda_u \mathcal{A}_{ij}}{\lambda_{ij}}\right)^k \mathsf{S}(n,k) \mathbb{E}\left[C^{k+1}(1)\right].$$

∎

## APPENDIX C

*Proof of Lemma 4:* If $Y_{ij}$ denotes the distance between the typical user and the tagged AP in $(i,j)$, then the distribution of $Y_{ij}$ is the distribution of $Z_{ij}$ conditioned on the user being associated with $(i,j)$. Therefore

$$\mathbb{P}(Y_{ij} > y) = \mathbb{P}\left(Z_{ij} > y | \text{ user is associated with } (i,j)\right) \quad (46)$$

$$= \frac{\mathbb{P}\left(Z_{ij} > y, \text{user is associated with } (i,j)\right)}{\mathbb{P}\left(\text{user is associated with } (i,j)\right)}. \quad (47)$$

Now using Lemma 2

$$\begin{aligned} &\mathbb{P}\left(Z_{ij} > y, \text{ user is associated with } (i,j)\right) \\ &= 2\pi\lambda_{ij} \int_{z>y} z \exp\left(-\pi \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) z^{2/\hat{\alpha}_k}\right) \mathrm{d}z. \end{aligned} \quad (48)$$

Using (47) and (48) we get

$$\begin{aligned} &\mathbb{P}(Y_{ij} > y) \\ &= \frac{2\pi\lambda_{ij}}{\mathcal{A}_{ij}} \int_{z>y} z \exp\left(-\pi \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) z^{2/\hat{\alpha}_k}\right) \mathrm{d}z, \end{aligned} \quad (49)$$

which leads to the PDF of $Y_{ij}$

$$f_{Y_{ij}}(y) = \frac{2\pi\lambda_{ij}}{\mathcal{A}_{ij}} y \exp\left(-\pi \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k) y^{2/\hat{\alpha}_k}\right). \quad (50)$$

∎

## APPENDIX D

*Proof of Lemma 5:* The SINR coverage of a user associated with an AP of $(i,j)$ is

$$\mathcal{S}_{ij}(\tau_{ij}) = \int_{y>0} \mathbb{P}(\mathtt{SINR}_{ij}(y) > \tau_{ij}) f_{Y_{ij}}(y) \mathrm{d}y. \tag{51}$$

Now $\mathbb{P}(\mathtt{SINR}_{ij}(y) > \tau_{ij})$ can be written as

$$\mathbb{P}\left(\frac{\mathrm{P}_{ij} h_y y^{-\alpha_j}}{\sum_{k\in\mathcal{V}_i} I_{ik} + \sigma_i^2} > \tau_{ij}\right) \tag{52}$$

$$= \mathbb{P}\left(h_y > y^{\alpha_j} {\mathrm{P}_{ij}}^{-1} \tau_{ij} \left\{\sum_{k\in\mathcal{V}_i} I_{ik} + \sigma_i^2\right\}\right) \tag{53}$$

$$= \mathbb{E}\left[\exp\left(-y^{\alpha_j}\tau_{ij}\mathrm{P}_{ij}^{-1}\left\{\sum_{k\in\mathcal{V}_i} I_{ik} + \sigma_i^2\right\}\right)\right] \tag{54}$$

$$\overset{(a)}{=} \exp\left(-\frac{\tau_{ij}}{\mathtt{SNR}_{ij}(y)}\right) \prod_{k\in\mathcal{V}_i} \mathbb{E}_{I_{ik}}\left[\exp\left(-y^{\alpha_j}\tau_{ij}\mathrm{P}_{ij}^{-1} I_{ik}\right)\right] \tag{55}$$

$$= \exp\left(-\frac{\tau_{ij}}{\mathtt{SNR}_{ij}(y)}\right) \prod_{k\in\mathcal{V}_i} \mathsf{M}_{I_{ik}}\left(y^{\alpha_j}\tau_{ij}\mathrm{P}_{ij}^{-1}\right), \tag{56}$$

where $\mathtt{SNR}_{ij}(y) = \frac{\mathrm{P}_{ij} y^{-\alpha_j}}{\sigma_i^2}$ and (a) follows from the independence of $I_{ik}$ and $\mathsf{M}_{I_{ik}}(s)$ is the the moment-generating function (MGF) of the interference. Expanding the interference term, the MGF of interference is given by

$$\mathsf{M}_{I_{ik}}(s) = \mathbb{E}_{\Phi_{ik},\Phi_{ik'},h_x,h_{x'}}\left[\exp\left(-s\mathrm{P}_{ik}\left\{\sum_{x\in\Phi_{ik}\setminus o} h_x x^{-\alpha_k} + \sum_{x'\in\Phi_{ik'}} h_{x'} x'^{-\alpha_k}\right\}\right)\right] \tag{57}$$

$$\overset{(a)}{=} \mathbb{E}_{\Phi_{ik}}\left[\prod_{x\in\Phi_{ik}\setminus o} \mathsf{M}_{h_x}\left(s\mathrm{P}_{ik}x^{-\alpha_k}\right)\right] \times \mathbb{E}_{\Phi_{ik'}}\left[\prod_{x'\in\Phi_{ik'}} \mathsf{M}_{h_{x'}}\left(s\mathrm{P}_{ik}x'^{-\alpha_k}\right)\right] \tag{58}$$

$$\overset{(b)}{=} \exp\left(-2\pi\lambda_{ik}\int_{z_{ik}}^{\infty}\left\{1-\mathsf{M}_{h_x}\left(s\mathrm{P}_{ik}x^{-\alpha_k}\right)\right\}x\mathrm{d}x\right) \times \exp\left(-2\pi\lambda_{ik'}\int_0^{\infty}\left\{1-\mathsf{M}_{h_{x'}}\left(s\mathrm{P}_{ik}x'^{-\alpha_k}\right)\right\}x'\mathrm{d}x'\right) \tag{59}$$

$$\overset{(c)}{=} \exp\left(-2\pi\lambda_{ik}\int_{z_{ik}}^{\infty}\frac{x}{1+(s\mathrm{P}_{ik})^{-1}x^{\alpha_k}}\mathrm{d}x - 2\pi\lambda_{ik'}\int_0^{\infty}\frac{x'}{1+(s\mathrm{P}_{ik})^{-1}x'^{\alpha_k}}\mathrm{d}x'\right), \tag{60}$$

where (a) follows from the independence of $\Phi_{ik}$,$\Phi_{ik'}$, $h_x$ and $h'_x$, (b) is obtained using the PGFL [19] of $\Phi_{ik}$ and $\Phi_{ik'}$, and (c) follows by using the MGF of an exponential RV with unit mean. In the above expressions, $z_{ik}$ is the lower bound on distance of the closest open access interferer in $(i,k)$ which can be obtained by using (1)

$$\mathrm{T}_{ij} y^{-\alpha_j} = \mathrm{T}_{ik} z_{ik}^{-\alpha_k} \text{ or } z_{ik} = (\hat{\mathrm{T}}_{ik})^{1/\alpha_k} y^{1/\hat{\alpha}_k}. \tag{61}$$

Using change of variables with $t = (s\mathrm{P}_{ik})^{-2/\alpha_k}x^2$, the integrals can be simplified as

$$\int_{z_{ik}}^{\infty}\frac{2x}{1+(s\mathrm{P}_{ik})^{-1}x^{\alpha_k}}\mathrm{d}x = (s\mathrm{P}_{ik})^{2/\alpha_k}\int_{(s\mathrm{P}_{ik})^{-2/\alpha_k}z_{ik}^2}^{\infty}\frac{\mathrm{d}t}{1+t^{\alpha_k/2}} = \mathsf{Z}\left(s\mathrm{P}_{ik},\alpha_k,z_{ik}^{\alpha_k}\right), \tag{62}$$

and

$$\int_0^{\infty}\frac{2x}{1+(s\mathrm{P}_{ik})^{-1}x^{\alpha_k}}\mathrm{d}x = \mathsf{Z}\left(s\mathrm{P}_{ik},\alpha_k,0\right), \tag{63}$$

where

$$\mathsf{Z}(a,b,c) = a^{2/b}\int_{(\frac{c}{a})^{2/b}}^{\infty}\frac{\mathrm{d}u}{1+u^{b/2}}.$$

This gives the MGF of interference

$$\mathsf{M}_{I_{ik}}(s) = \exp\left(-\pi(s\mathrm{P}_{ik})^{2/\alpha_k} \times\left\{\lambda_{ik}\mathsf{Z}\left(1,\alpha_k,\frac{z_{ik}^{\alpha_k}}{s\mathrm{P}_{ik}}\right)+\lambda_{ik'}\mathsf{Z}\left(1,\alpha_k,0\right)\right\}\right). \tag{64}$$

Using $s = y^{\alpha_j}\tau_{ij}\mathrm{P}_{ij}^{-1}$ with $z_{ik}$ from (61) for MGF of interference in (56) we get

$$\mathbb{P}(\mathtt{SINR}_{ij}(y) > \tau_{ij}) = \exp\left(-\frac{\tau_{ij}}{\mathtt{SNR}_{ij}(y)} - \pi\sum_{k\in V_i} y^{2/\hat{\alpha}_k} D_{ij}\left(k,\tau_{ij}\right)\right), \tag{65}$$

where

$$D_{ij}(k,\tau_{ij}) = \hat{\mathrm{P}}_{ik}^{2/\alpha_k}\left\{\lambda_{ik}\mathsf{Z}\left(\tau_{ij},\alpha_k,\hat{\mathrm{P}}_{ik}^{-1}\hat{\mathrm{T}}_{ik}\right)+\lambda_{ik'}\mathsf{Z}\left(\tau_{ij},\alpha_k,0\right)\right\}.$$

Using (51) along with Lemma 4 gives

$$\mathcal{S}_{ij}(\tau_{ij}) = \frac{2\pi\lambda_{ij}}{\mathcal{A}_{ij}}\int_{y>0} y\exp\left(-\frac{\tau_{ij}}{\mathtt{SNR}_{ij}(y)} - \pi\left\{\sum_{k\in\mathcal{V}_i} D_{ij}(k,\tau_{ij})y^{2/\hat{\alpha}_k} + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)y^{2/\hat{\alpha}_k}\right\}\right)\mathrm{d}y. \tag{66}$$

Using the law of total probability we get

$$\mathcal{S} = \sum_{(i,j)\in\mathcal{V}^o}\mathcal{S}_{ij}(\tau_{ij})\mathcal{A}_{ij}, \tag{67}$$

which gives the overall SINR coverage of a typical user. ■

## APPENDIX E

*Proof of Proposition 1:* In the described setting SIR coverage can be written as

$$\mathcal{S} = \sum_{(i,j)\in\mathcal{V}^o}\frac{\lambda_{ij}}{\sum_{k\in\mathcal{V}_i} D_{ij}(k,\tau_{ij}) + \sum_{(m,k)\in\mathcal{V}^o} G_{ij}(m,k)}, \tag{68}$$

and with $\mathcal{V} = \{(1,q),(2,r)\}$, $\lambda_{2r} = a\lambda_{1q}$, and $\mathrm{B}_{2r} = b\mathrm{B}_{1q}$

$$\begin{aligned}\mathcal{S} &= \frac{\lambda_{1q}}{\lambda_{1q}\mathsf{Z}(\tau_{1q},\alpha,1)+\lambda_{1q}+\lambda_{2r}(\hat{\mathrm{P}}_{2r}\hat{\mathrm{B}}_{2r})^{2/\alpha}} \\ &\quad + \frac{\lambda_{2r}}{\lambda_{2r}\mathsf{Z}(\tau_{2r},\alpha,1)+\lambda_{2r}+\lambda_{1q}(\hat{\mathrm{P}}_{1q}\hat{\mathrm{B}}_{1q})^{2/\alpha}} \\ &= \frac{1}{\mathsf{Z}(\tau_{1q},\alpha,1)+1+a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha}} \\ &\quad + \frac{1}{\mathsf{Z}(\tau_{2r},\alpha,1)+1+\frac{1}{a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha}}}.\end{aligned}$$

The gradient of $\mathcal{S}$ with respect to association bias $\nabla_b\mathcal{S}$ is zero at

$$\begin{aligned}b_{\mathrm{opt}} &= \arg\max_b \left\{ \left(\mathsf{Z}(\tau_{1q},\alpha,1)+1+a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha}\right)^{-1} \right. \\ &\quad \left. + \left(\mathsf{Z}(\tau_{2r},\alpha,1)+1+\frac{1}{a(\hat{\mathrm{P}}_{2r}b)^{2/\alpha}}\right)^{-1} \right\} \\ &= \frac{\mathrm{P}_{1q}}{\mathrm{P}_{2r}}\left(\frac{\mathsf{Z}(\tau_{1q},\alpha,1)}{a\mathsf{Z}(\tau_{2r},\alpha,1)}\right)^{\alpha/2}.\end{aligned}$$

With algebraic manipulation, it can be shown that for all $b > b_{\mathrm{opt}}$ $\nabla_b\mathcal{S} < 0$ and for all $b < b_{\mathrm{opt}}$ $\nabla_b\mathcal{S} > 0$ and hence $\mathcal{S}$ is strictly quasiconcave in $b$ and $b_{\mathrm{opt}}$ is the unique mode. Using Lemma 2, the optimal traffic offload fraction is obtained as

$$\begin{aligned}\mathcal{A}_2 &= \frac{\lambda_{2r}}{G_{2r}(r)} = a\left\{a+\left(\frac{\mathrm{P}_{1q}}{\mathrm{P}_{2r}b_{\mathrm{opt}}}\right)^{2/\alpha}\right\}^{-1} \\ &= \frac{\mathsf{Z}(\tau_{1q},\alpha,1)}{\mathsf{Z}(\tau_{2r},\alpha,1)+\mathsf{Z}(\tau_{1q},\alpha,1)}. \qquad (69)\end{aligned}$$

The corresponding SIR coverage can then obtained by substituting the optimal bias value in (68). ■

## REFERENCES


[1] S. Singh, H. S. Dhillon, and J. G. Andrews, "Downlink rate distribution in multi-RAT heterogeneous networks," in *2013 IEEE ICC*.

[2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10–21, June 2011.

[3] Qualcomm, "A comparison of LTE-Advanced HetNets and WiFi," white paper. Available: http://goo.gl/BFMFR, Sept. 2011.

[4] Juniper, "WiFi and femtocell integration strategies 2011-2015," white paper. Available: http://www.juniperresearch.com/, Mar. 2011.

[5] Qualcomm, "A 3G/LTE Wi-Fi Offload Framework," white paper. Available: http://goo.gl/91EqQ, June 2011.

[6] S. Singh, J. G. Andrews, and G. de Veciana, "Interference shaping for improved quality of experience for real-time video streaming," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1259–1269, Aug. 2012.

[7] L. Wang and G. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networks–a tutorial," *IEEE Commun. Surveys & Tutorials*, vol. PP, no. 99, pp. 1–22, 2012.

[8] E. Stevens-Navarro, Y. Lin, and V. Wong, "An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1243–1254, Mar. 2008.

[9] K. Premkumar and A. Kumar, "Optimum association of mobile wireless devices with a WLAN-3G access network," in *Proc. 2006 IEEE ICC*, pp. 2002–2008.

[10] D. Kumar, E. Altman, and J.-M. Kelif, "Globally optimal user-network association in an 802.11 WLAN and 3G UMTS hybrid cell," in *2007 ITC*.

[11] ——, "User-network association in an 802.11 WLAN & 3G UMTS hybrid cell: individual optimality," in *2007 IEEE Sarnoff Symp.*

[12] K. Khawam, M. Ibrahim, J. Cohen, S. Lahoud, and S. Tohme, "Individual vs. global radio resource management in a hybrid broadband network," in *2011 IEEE ICC*.

[13] S. Elayoubi, E. Altman, M. Haddad, and Z. Altman, "A hybrid decision approach for the association problem in heterogeneous networks," in *Proc. 2010 IEEE INFOCOM*, pp. 1–5.

[14] F. Moety, M. Ibrahim, S. Lahoud, and K. Khawam, "Distributed heuristic algorithms for RAT selection in wireless heterogeneous networks," in *2012 IEEE WCNC*.

[15] T. Togo, I. Yoshii, and R. Kohno, "Dynamic cell-size control according to geographical mobile distribution in a DS/CDMA cellular system," in *1998 IEEE PIMRC*, vol. 2, pp. 677–681.

[16] A. Jalali, "On cell breathing in CDMA networks," in *Proc. 1998 IEEE ICC*, vol. 2, pp. 985–988.

[17] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," in *Proc. 2004 ACM MobiCom*, pp. 302–314.

[18] Nokia Siemens Networks, Nokia, "Aspects of pico node range extension," 3GPP TSG RAN WG1 Meeting 61, R1-103824, 2010. Available: http://goo.gl/XDKXI

[19] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*. John Wiley & Sons, 1996.

[20] S. Singh, O. Oyman, A. Papathanassiou, D. Chatterjee, and J. G. Andrews, "Video capacity and QoE enhancements over LTE," in *2012 IEEE ICC Workshop Realizing Advanced Video Optimized Wireless Netw.*

[21] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of $K$-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[22] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: a comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[23] S. Mukherjee, "Distribution of downlink SINR in heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 575–585, Apr. 2012.

[24] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[25] B. Blaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using Poisson processes to model lattice cellular networks," in *2013 IEEE INFOCOM*. Available: http://arxiv.org/abs/1207.7208

[26] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks." Available: http://arxiv.org/abs/1109.2992

[27] D. Cao, S. Zhou, and Z. Niu, "Optimal base station density for energy-efficient heterogeneous cellular networks," in *Proc. 2012 IEEE ICC*, pp. 4379–4383.

[28] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Load-aware modeling and analysis of heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, to appear. Available: http://arxiv.org/abs/1204.1091

[29] X. Lin, J. G. Andrews, and A. Ghosh, "Modeling, analysis and design for carrier aggregation in heterogeneous cellular networks," *IEEE Trans. Commun.*, submitted, [Online]. Available: http://arxiv.org/abs/1211.4041

[30] E. N. Gilbert, "Random subdivisions of space into crystals," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 958–972, Sept. 1962. Available: http://www.jstor.org/stable/2237872

[31] A. Hinde and R. Miles, "Monte Carlo estimates of the distributions of the random polygons of the Voronoi tessellation with respect to a Poisson process," *J. Statistical Computation Simulation*, vol. 10, no. 3-4, pp. 205–223, 1980. Available: http://www.tandfonline.com/doi/abs/10.1080/00949658008810370

[32] J.-S. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518–526, Nov. 2007. Available: http://www.sciencedirect.com/science/article/pii/S0378437107007546

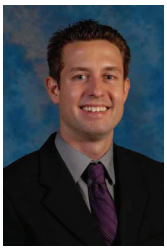
[33] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "Stochastic analysis of spatial and opportunistic Aloha," *IEEE J. Sel. Areas Commun.*, pp. 1105–1119, Sept. 2009.

[34] P. F. Ash and E. D. Bolker, "Generalized Dirichlet tessellations," *Geometriae Dedicata*, vol. 20, pp. 209–243, 1986. Available: http://dx.doi.org/10.1007/BF00164401

[35] Google, "Google WiFi Mountain View coverage map." Available: http://wifi.google.com/city/mv/apmap.html.

**Sarabjot Singh** (S'09) received his Bachelor of Technology (B.Tech.) in Electronics and Communication Engineering from Indian Institute of Technology, Guwahati in 2010 and was awarded the prestigious President of India Gold Medal. He is currently working towards his Ph.D. in Electrical Engineering at the University of Texas at Austin. His research interests broadly span various aspects of wireless communication and networks with current focus on design and analysis of radio resource management for video aware wireless systems, which includes interplay of traffic offloading, scheduling, and interference coordination in wireless networks. He has held internship positions at Intel Corporation in Santa Clara, CA and Qualcomm Inc. in San Diego, CA.

**Harpreet S. Dhillon** (S'11) received the B.Tech. degree in Electronics and Communication Engineering from IIT Guwahati, India, in 2008 and the M.S. in Electrical Engineering from Virginia Tech in 2010. He is currently a Ph.D. student at The University of Texas at Austin, where his research has focused on the modeling and analysis of heterogeneous cellular networks using tools from stochastic geometry, point process theory and spatial statistics. His other research interests include interference channels, multiuser MIMO systems and cognitive radio networks. He is the recipient of the Microelectronics and Computer Development (MCD) fellowship from UT Austin and was also awarded the Agilent Engineering and Technology Award 2008. He has held summer internships at Alcatel-Lucent Bell Labs in Crawford Hill, NJ, Samsung Dallas Technology Lab in Richardson, TX, Qualcomm Inc. in San Diego, CA, and Cercom, Politecnico di Torino in Italy.



**Jeffrey Andrews** (S'98, M'02, SM'06, F'13) received the B.S. in Engineering with High Distinction from Harvey Mudd College in 1995, and the M.S. and Ph.D. in Electrical Engineering from Stanford University in 1999 and 2002, respectively. He is a Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin, where he was the Director of the Wireless Networking and Communications Group (WNCG) from 2008-12. He developed Code Division Multiple Access systems at Qualcomm from 1995-97, and has consulted for entities including Verizon, the WiMAX Forum, Intel, Microsoft, Apple, Samsung, Clearwire, Sprint, and NASA. He is a member of the Technical Advisory Boards of Accelera and Fastback Networks.

Dr. Andrews is co-author of two books, Fundamentals of WiMAX (Prentice-Hall, 2007) and Fundamentals of LTE (Prentice-Hall, 2010), and holds the Earl and Margaret Brasfield Endowed Fellowship in Engineering at UT Austin, where he received the ECE department's first annual High Gain award for excellence in research. He is a Distinguished Lecturer for the IEEE Vehicular Technology Society, served as an associate editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2004-08, was the Chair of the 2010 IEEE Communication Theory Workshop, and is the Technical Program co-Chair of ICC 2012 (Comm. Theory Symposium) and Globecom 2014. He is an elected member of the Board of Governors of the IEEE Information Theory Society and an IEEE Fellow.

Dr. Andrews received the National Science Foundation CAREER award in 2007 and has been co-author of five best paper award recipients, two at Globecom (2006 and 2009), Asilomar (2008), the 2010 IEEE Communications Society Best Tutorial Paper Award, and the 2011 Communications Society Heinrich Hertz Prize. His research interests are in communication theory, information theory, and stochastic geometry applied to wireless cellular and ad hoc networks.